Search for di-Higgs production to the $b\bar{b}\tau^+\tau^-$ final state using Support Vector Machines

with the ATLAS detector at the Large Hadron Collider in pp collisions at $\sqrt{s} = 13\,TeV$

Thomas James Stevenson
Particle Physics Research Centre
Queen Mary University of London

Submitted in partial fulfillment of the requirements of the Degree of Doctor of Philosophy

I, Thomas James Stevenson, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

Signature:

Date:

Details of collaboration and publications:

This work was supported by the Science and Technology Facilities Council

I would first like to thank Dr Adrian Bevan and Dr Jon Hays for their supervision and regular discussion throughout my PhD. Their guidance has been invaluable to my development as a physicist, offering incites from their knowledge of all aspects of research whilst allowing me the freedom to explore novel ideas.

I am extremely grateful to Dr Agni Bethani who has worked closely with me and been a mentor from the beginning of our analysis. I would also like to show my appreciation to the ATLAS bb$\tau\tau$ analysis group. We worked hard together through many challenging issues to achieve, at present, the worlds best di-Higgs result and I wish them luck for the continuing analysis in the future.

It has been a pleasure to be part of PPRC and SPA and I wish to thank everyone I have known and worked with there. My time at Queen Mary has been amazing and something I will never forget.

While at CERN I was fortunate to be surrounded by many marvellous and interesting people. In particular I would like to thank my friends Abby, Andy, Andy, James, Jon, Laurie and Tom for all the great times and much wine we shared together.

I would like to thank Professor Becky Parker and all my friends from the Langton Star Centre for inspiring me to pursue a degree and doctorate in physics and for our many trips to CERN.

To my family, thanks for their support and belief in me, and their provision of plentiful food and drink on my far too infrequent visits home.

Finally I would like to thank my wonderful partner Sarah for supporting me and keeping me sane over these last three and a half years. I appreciate her patience with my occasional late nights and weekends working to meet deadlines and the time I spent abroad.
Without your unwavering support this would have not been possible.

Abstract

This thesis presents a search for resonant and standard model non-resonant di-Higgs production in the $b\bar{b}\tau^+\tau^-$ decay channel, with the semi-leptonic di-tau state. The search uses data collected by the ATLAS experiment at the LHC of pp-collisions at $\sqrt{s} = 13\,\text{TeV}$ during 2015 and 2016, corresponding to an instantaneous luminosity of $36.1\,\text{fb}^{-1}$. No deviation from the Standard Model prediction is observed. For the non-resonant standard model di-Higgs production an upper limit is set on the cross-section compared to the Standard Model prediction, $\frac{\sigma}{\sigma_{\text{SM}}} < 24.44$ @ 95%CL. Upper limits are set on the cross-section times branching ratio for the resonant searches, where the results are interpreted in terms of constraints on a 2HDM heavy scalar Higgs model and a Randall-Sundrum Kaluza-Klein graviton model. The use of Support Vector Machines is investigated and benchmarked against the nominal results of the Boosted Decision Tree multivariate analysis, in pursuit of potential future improvements.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The discovery of a new particle in 2012 at the Large Hadron Collider (LHC) by the ATLAS and CMS experiments [1][2] consistent with the Standard Model (SM) Higgs boson [3][4], h, marked a significant step forward in improving our understanding of the electroweak symmetry breaking mechanism (EWSB) in the SM. The measurement of the Higgs boson mass of $m_h = 125.09 \pm 0.21(\text{stat}) \pm 0.11(\text{syst})\,\text{GeV}$ [5] constrains the structure of the Higgs scalar potential and the Higgs boson self-coupling, and measuring the latter provides an independent test to verify EWSB in the SM and the shape of the potential.

In the SM Higgs bosons can be pair produced via the Higgs self-interaction or top quark loop. The measurement of the cross-section of this process would allow the extraction of the trilinear Higgs self-coupling (λ_{hhh}) and is one of the primary goals for the High Luminosity LHC (HL-LHC). Currently at the LHC di-Higgs production mainly occurs via the gluon-gluon fusion production mechanism with the two leading processes destructively interfering leading to a very small SM cross-section prediction of $\sigma_{\text{NNLL}} = 33.49^{+1.44}_{-2.00}(\text{scale}) \pm 1.67(\text{theory}) \pm 0.77(\alpha_s) \pm 0.70(\text{PDF})\,\text{fb}$ [6], three orders of magnitude less than other Higgs processes.

Several extensions beyond the Standard Model (BSM) can enhance the predicted cross-section for di-Higgs production. Modifications to λ_{hhh} and the top Yukawa coupling can lead to significant enhancements. Other theories that predict the decay of heavy resonances to two SM $m_h = 125.09\,\text{GeV}$ Higgs bosons, such as a spin-2 Kaluza-Klein (KK) graviton, G^*_{KK}, in the bulk Randal-Sundrum (RS) [7][8] model or a heavy spin-0 Higgs, H, in two Higgs doublet models (2HDM) [9][10], could also give enhancements.

In this thesis the search for BSM resonant and SM non-resonant Higgs pair production decaying to a final state with two b-jets and two τ leptons, one decaying leptonically

and the other hadronically, is described. The di-Higgs to $bb\tau\tau$[1] process has the third highest branching fraction (7.3%), however this has a relatively clean signature and better separation from the large multi-jet and $t\bar{t}$ backgrounds compared to the $bbbb$ and $bbWW$ channels with higher branching fractions. The analysis is performed using data collected by the ATLAS detector during 2015 and 2016 corresponding to an integrated luminosity of 36.1fb^{-1}.

The nominal analysis utilises Boosted Decision Trees (BDTs) to better separate the signal from the backgrounds and improve sensitivity to the various signal models. The use of Support Vector Machines (SVMs), an alternative machine learning algorithm, as well as other multivariate analysis (MVA) techniques are explored and benchmarked against the BDT analysis.

Previously during Run 1 of the LHC searches for di-Higgs production were performed at the LHC by the ATLAS experiment [11] using the data collected in 2012 with a centre-of-mass energy of $\sqrt{s} = 8\,\text{TeV}$. For the $bb\tau\tau$ process only the $\tau_{\text{lep}}\tau_{\text{had}}$ decay channel was considered and a cut based analysis was used. Upper limits were set for BSM resonances, X, on the production cross-section times branching ratio to di-Higgs with masses between $260\,\text{GeV} < m_X < 1\,\text{TeV}$. An upper limit of 1.6 pb was set on SM non-resonant di-Higgs production, 160 times the standard model prediction. This result was combined with other ATLAS analyses of the $bbbb$ [12], $bb\gamma\gamma$ [13] and $bbWW^*$ [11] yielding an upper limit on the SM non-resonant di-Higgs production cross-section of 0.67 pb, 70 times the predicted standard model cross-section. A Run 2 search for Higgs boson pair production to the $bb\tau\tau$ final state by the CMS experiment using $35.9\,\text{fb}^{-1}$ of data from 2015 and 2016 at $\sqrt{s} = 13\,\text{TeV}$ [14] excluded non-resonant di-Higgs production for cross-sections above 28 times the SM prediction.

The work contained in this thesis was conducted as part of a wider analysis team working on the search for di-Higgs to $bb\tau\tau$ on ATLAS, including also the final state where both taus decay hadronically, which will be published soon. The contributions to the analysis of the author were focused on the multi-variate analysis optimisation (particularly for the lepton-plus-tau trigger channel), work on developing the fake tau background estimation and the systematics related to this, as well as the combination of the single lepton trigger and lepton-plus-tau trigger results. Contributions to these and other areas were obtained from other members of the analysis team for which the author is grateful.

[1] $bb\tau\tau$ is used as shorthand to refer to the process where one Higgs decays to a b-quark and an anti-b-quark, $h \to b\bar{b}$ and the other Higgs decays to a tau and an anti-tau, $h \to \tau^+\tau^-$.

This thesis is structured as follows. Firstly an overview of the Standard Model including the Brout-Englert-Higgs mechanism as well as a summary of Higgs production at the LHC are presented in Chapter 2. The machine learning methods used in the analysis are described in Chapter 3. Chapter 4 gives a summary of the LHC and the ATLAS detector. The Monte Carlo simulations of the signal and background processes that contribute are discussed in Chapter 5. The definitions of the reconstructed physical objects are detailed in Chapter 6 and the event selection for the analysis is described in Chapter 7. The strategy for the estimation of backgrounds is summarised in Chapter 8. Chapter 9 details the systematic uncertainties considered. A description of the statistical procedure used and the results of the analysis are presented in Chapter 10. The use of alternative multivariate techniques are explored in Chapter 11. Finally a summary of the analysis and conclusions are given in Chapter 12.

Chapter 2

Physics Theory

The relevant theoretical background is outlined in this chapeter. The particle content and interactions of the SM are briefly described, with a particular focus on spontaneous symmetry breaking and how it relates to the Higgs boson. The production mechanisms for Higgs bosons at the LHC are discussed as well as both SM and BSM pair production of the Higgs and its decays, which can result in the $bb\tau\tau$ final state. The main backgrounds which mimic the final state are also briefly covered. More detailed discussion of the topics and formalism contained within this chapter can be found in References [10], [15] & [16].

2.1 The Standard Model

Developed throughout the latter half of the 20th century, the SM is the theory describing the fundamental particles and their short-range interactions via three of the four fundamental forces; the electromagnetic, weak and strong forces. It is based on the product of the groups $SU(3)_c \times SU(2)_L \times U(1)_Y$ and the mathematical framework of quantum field theory, with $SU(3)_c$ describing quantum chromodynamics which parametrises the strong force, while $SU(2)_L \times U(1)_Y$ parametrises the combination of electromagnetism and weak forces via electroweak interactions. The recent discovery of the Higgs boson [1][2] responsible for the mass of the particles in the SM completes this picture.

2.1.1 Particles

The particle content of the SM can be split into two distinct categories; particles which have half-integer spin obeying Fermi-Dirac statistics [17] referred to as fermions, and particles with integer spin that obey Bose-Einstein statistics called bosons. Due to these spin-statistics the fermions are subject to the Pauli exclusion principle [18], meaning that in a system no two identical fermions can occupy the same quantum state, while bosons do not follow this.

2.1.1.1 Fermions

There are 12 elementary fermions in the SM which can be further split into groups depending on the charges they carry and the particles they interact with, summarised in Table 2.1. These are also grouped into pairs of particles which exhibit similar properties forming three generations.

The first set of particles are referred to as quarks with the six different flavours, up & down, charm & strange, top & bottom, where each pair belongs to a generation, increasing in mass. The up-type quark of each pair (u,c,t) have electric charge of $+\frac{2}{3}e$, where e is the fundamental electric charge, while the down-type quarks (d, s, b) have $-\frac{1}{3}e$. The main property distinguishing quarks is colour charge, a quantum number parametrising their interaction with the strong force which has the value red (r), green (g) or blue (b). As a consequence quarks are very strongly bound to each other and as such are not found individually in nature, a phenomenon referred to as colour confinement, but rather in colourless[1] bound states of three (anti-) quarks referred to as (anti-) baryons or in quark anti-quark pairs called mesons.

Generation	Quark	Symbol	Charge Q	Colour	Mass [MeV]
1	up	u	$+\frac{2}{3}$	r, g, b	2.2
	down	d	$-\frac{1}{3}$		4.7
2	charm	c	$+\frac{2}{3}$	r, g, b	1280
	strange	s	$-\frac{1}{3}$		96
3	top	t	$+\frac{2}{3}$	r, g, b	173100
	bottom	b	$-\frac{1}{3}$		4180
Generation	Lepton	Symbol	Charge Q	Colour	Mass [MeV]
1	electron	e	-1	—	0.511
	electron neutrino	ν_e	0		< 0.000002
2	muon	μ	-1	—	105.7
	muon neutrino	ν_μ	0		< 0.000002
3	tau	τ	-1	—	1776.86
	tau neutrino	ν_τ	0		< 0.000002

Table 2.1: The elementary fermions of the Standard Model. The names and symbols of each particle with the generation they belong to, their electric charge (Q), colour charge and mass detailed. [19]

The second group of 6 fermions are leptons. These do not interact with the strong force

[1] Colourless states are those with an equal mixture of red, green and blue (rgb), anti-red, anti-green and anti-blue $(\overline{r}\,\overline{g}\,\overline{b})$ or a colour anti-colour pair ($r\overline{r}, g\overline{g}, b\overline{b}$).

as they do not possess colour charge. These again are split into three generations, with each containing a lepton with electric charge of $-1e$, the electron, muon and tau, and a corresponding neutrino with zero electric charge. In the SM theory the neutrinos have no intrinsic mass, however experimental observation of neutrino mixing implies they have a small non-zero mass, constrained to be $< 2\,\text{eV}$ [19].

2.1.1.2 Bosons

In the SM the gauge or vector (spin-1) bosons mediate the interactions between particles. Their properties are summarised in Table 2.2. Firstly there is the photon, a massless particle which mediates the electromagnetic force between electrically charged particles, though they themselves have zero electric charge. The gluons also have no mass and are responsible for the strong interactions coupling to colour charged particles, interacting with the quarks and themselves as they also possess colour charge. There are eight different gluons labeled with various combinations of colour and anti-colour. There are three massive gauge bosons, the W^+, W^- and Z^0, which mediate the weak interactions between the different flavours of leptons and quarks. The $W^\pm$ are electrically charged and interact also via electromagnetic interactions, while the electrically neutral Z^0 does not. Finally there is a massive scalar (spin-0) Higgs boson which, as will be seen later, is responsible for giving mass to the gauge bosons and itself via spontaneous symmetry breaking and to fermions through Yukawa interactions.

Boson	Interaction	Charge Q	Colour	Spin	Mass [MeV]
γ	EM	0	—	1	0
$W^\pm$	Weak	±1	—	1	80385
Z^0	Weak	0	—	1	91188
g	Strong	0	8 combinations	1	0
h	Yukawa	0	—	0	125090

Table 2.2: The Bosons of the Standard Model. The interaction they mediate, their electric charge (Q), colour, spin and mass detailed. [19]

2.1.2 Quantum Chromodynamics

Quantum Chromodynamics (QCD) is the gauge field theory parametrised by the non-abelian group $SU(3)_C$ describing the interactions of the strong force between coloured particles, quarks and gluons, in the SM. Quarks are represented by and transform as colour

triplets in the fundamental representation of $SU(3)$

$$\psi_f(x) = \begin{pmatrix} \psi_f^r(x) \\ \psi_f^g(x) \\ \psi_f^b(x) \end{pmatrix}, \tag{2.1}$$

where f is the flavour index and r, g, b are the indices corresponding to the colour quantum numbers. The gluons are associated with the eight gauge fields G_a^μ with the Lagrangian given by

$$\mathscr{L}_{QCD} = -\frac{1}{4} G_a^{\mu\nu} G_{\mu\nu}^a + \sum_f \left[\overline{\psi}_f \left(i\gamma^\mu \partial_\mu - m \right) \psi_f + g_s \left(\overline{\psi} \gamma^\mu T_a \psi \right) G_\mu^a \right] \tag{2.2}$$

where $T_a = \lambda_a/2$ the eight generators of $SU(3)$ and λ_a are the Gell-Mann matrices [20]. The field strength tensor corresponding to the dynamics of the gluons is

$$G_a^{\mu\nu} = \partial^\mu G_a^\nu - \partial^\nu G_a^\mu + g_s f_{abc} G_b^\mu G_c^\nu. \tag{2.3}$$

Here f_{abc} are the totally antisymmetric $SU(3)$ structure constants and the generators obey the Lie algebra $[T_a, T_b] = i f_{abc} T_c$. The term containing these structure constants corresponds to the three and four-point gluon self-interaction vertices. This Lagrangian is invariant under local $SU(3)$ variations of the form

$$\psi_f(x) \to e^{iT^a \alpha_a(x)} \psi_f(x). \tag{2.4}$$

In addition QCD also requires the inclusion of terms for scalar anti-commuting fields, referred to as ghost particles, which cancel unphysical polarisation states of the gluons, as well as gauge fixing terms allowing inverse propagators to be defined.

2.1.3 The Glashow-Salam-Weinberg Model of Electroweak Interactions

The combination of weak and electromagnetic theories to form a unified structure, first proposed by Glashow [21], Salam [22] and Weinberg [23], electroweak (EW) interactions in the SM are described by the gauge groups $SU(2)_L \times U(1)_Y$. The subscript L here denotes that only the components of the fields with left-handed chirality[2] transform under $SU(2)_L$. Quantum Electrodynamics (QED) [16], the theory describing electromagnetic interactions, was understood to possess a gauge symmetry with the photon as the gauge boson. In order to extend this theory to other interactions additional gauge bosons are required to be introduced for the gauge symmetry to hold, as will be seen later.

[2]Chirality eigenstates for left- and right-handed states are obtained using, $P_L = \frac{1}{2}(1 - \gamma_5)$ and $P_R = \frac{1}{2}(1 + \gamma_5)$ respectively with $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$.

The fermions are described by weak isospin doublets, separately for leptons and quarks, containing left-handed particles as

$$\psi_L^{l=e,\mu,\tau} = \begin{pmatrix} \nu_l \\ l_l \end{pmatrix}_L, \quad \psi_L^{f=1,2,3} = \begin{pmatrix} u_f \\ d'_f \end{pmatrix}_L. \tag{2.5}$$

Here $d'_i = \sum_j V_{ij}^{CKM} d_j$ as the quark weak interaction eigenstates are observed to be mixtures of the flavour eigenstates, where V^{CKM} is the unitary Cabibbo-Kobayashi-Maskawa mixing matrix. The magnitudes of the elements of this matrix are measured experimentally [24] to be

$$V^{CKM} = \begin{pmatrix} |V_{ud}| \simeq 0.974 & |V_{us}| \simeq 0.225 & |V_{ub}| \simeq 0.004 \\ |V_{cd}| \simeq 0.225 & |V_{cs}| \simeq 0.973 & |V_{cb}| \simeq 0.043 \\ |V_{td}| \simeq 0.009 & |V_{ts}| \simeq 0.042 & |V_{tb}| \simeq 0.999 \end{pmatrix}. \tag{2.6}$$

The quantum number associated with weak interactions is weak isospin, T, with the third component T^3 being conserved. The left-handed fermion doublets have $T = \frac{1}{2}$, with the upper and lower components of the doublets having $T^3 = \pm\frac{1}{2}$ respectively. As the right-handed fermions do not couple to weak interactions, they behave as $SU(2)_L$ singlets and as such have both $T = 0$ and $T^3 = 0$. However both the left- and right-handed components interact under $U(1)_Y$, where the associated quantum number is weak hypercharge, Y_W, related to the third component of weak isospin by the electric charge, Q as

$$Q = T^3 + \frac{Y_W}{2}. \tag{2.7}$$

A crucial consequence of this relation is that all quantum numbers, Q, T, T^3 and Y_W, which would be associated to right-handed neutrino states are zero and as such are not present, fitting in with experimental observations. Finally the gauge bosons are introduced with a weak isospin triplet of gauge fields $W^\mu_{a=1,2,3}$ with coupling constant g to gauge the $SU(2)_L$ symmetry and $U(1)_Y$ related to the field B^μ with coupling constant $\frac{g'}{2}$. It will later be seen after spontaneous symmetry breaking that combinations of these correspond to the physical $W^\pm$ and Z^0 bosons and the photon.

The EW Lagrangian is written as

$$\begin{aligned} \mathscr{L}_{EW} = &-\frac{1}{4} W_a^{\mu\nu} W^a_{\mu\nu} - \frac{1}{4} B^{\mu\nu} B_{\mu\nu} \\ &+ \sum_\psi \left(\overline{\psi}_L \gamma^\mu \left[i\partial_\mu - \frac{1}{2} g \tau_a \cdot W^a_\mu - \frac{g'}{2} Y_W B_\mu \right] \psi_L + \overline{\psi}_R \gamma^\mu \left[i\partial_\mu - \frac{g'}{2} Y_W B_\mu \right] \psi_R \right), \end{aligned} \tag{2.8}$$

where $\tau_{a=1,2,3}$, the Pauli matrices, are the generators of $SU(2)$, satisfying the Lie algebra $[\tau_i, \tau_j] = 2i\varepsilon_{ijk}\tau_k$ where ε_{ijk} is the anti-symmetric Levi-Civita tensor. $\mathscr{L}_{EW}$ is invariant

under the $SU(2)_L$ and $U(1)_Y$ gauge transformations

$$\psi_L \to \exp\left[-ig\tau_a \cdot \theta^a(x) + i\frac{g'}{2}\chi(x)\right]\psi_L,$$
$$\psi_R \to \exp\left[ig'\chi(x)\right]\psi_R, \tag{2.9}$$

with the gauge fields transforming as $W_a^{'\mu} = W_a^\mu + g\varepsilon_{abc}\theta_b W_c^\mu + \partial^\mu\theta^a$ and $B^{'\mu} = B^\mu + \partial^\mu\chi$. For the kinetic energy terms of the bosons the field strength tensor for $SU(2)_L$ is given by $W_a^{\mu\nu} = \partial^\mu W_a^\nu - \partial^\nu W_a^\mu - g\varepsilon_{abc}W_b^\mu W_c^\nu$ while for $U(1)_Y$ the B^μ field has the Abelian field strength tensor $B^{\mu\nu} = \partial^\mu B^\nu - \partial^\nu B^\mu$.

Noticeable omissions from this Lagrangian are Dirac mass terms for the leptons and quarks, $m(\overline{\psi}_R\psi_L + \overline{\psi}_L\psi_R)$, as although these are gauge invariant under $U(1)_Y$ transformations, they break the gauge invariance under $SU(2)_L$ transformations. Furthermore this issue arises for the B^μ and W_a^μ fields, which cannot have mass terms of the form $\frac{1}{2}M_B^2 B^\mu B_\mu$ and $M_W^2 W_a^\mu W_\mu^a$ as these break the invariance of the $U(1)_Y$ and $SU(2)_L$ gauge transformations respectively. The solution to this comes from the previously undiscussed Higgs bosons and the Brout-Englert-Higgs mechanism described in the following.

2.1.4 Spontaneous Symmetry Breaking and The Brout-Englert-Higgs Mechanism

The solution to giving the $W^\pm$ and Z^0 bosons mass while leaving the photon massless and, as an additional consequence, providing mass terms for the fermions comes from the spontaneous symmetry breaking of an additional complex scalar field. Proposed in the 1960s by several groups who independently developed parts of the theory, Brout and Englert [25], Higgs [26][27] and Guralnik, Hagen and Kibble [28], it also predicts the existence of a massive scalar particle to mediate the field, the Higgs boson. The process of spontaneous symmetry breaking is discussed in the following, starting with a trivial example before finally moving to the symmetry breaking pattern of the SM.

2.1.4.1 Spontaneous Symmetry Breaking of a Global Gauge Symmetry

First a two component complex scalar field is considered

$$\phi \equiv \frac{1}{\sqrt{2}}(\phi_1 + i\phi_2),$$
$$\phi^\dagger \equiv \frac{1}{\sqrt{2}}(\phi_1 - i\phi_2), \tag{2.10}$$

described by the Lagrangian

$$\mathscr{L} = T - V = \partial_\mu \phi^\dagger \partial^\mu \phi - \mu^2 \phi^\dagger \phi + \lambda(\phi^\dagger \phi)^2, \tag{2.11}$$

which is globally gauge invariant under phase variations $\phi \to \exp^{i\chi} \phi$, where the scalar potential is

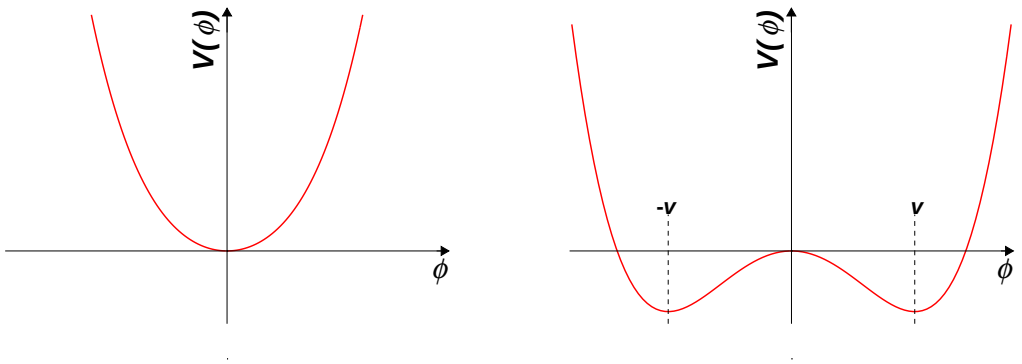
$$V(\phi) = \frac{1}{2}\mu^2(\phi_1^2 + \phi_2^2) - \frac{1}{4}\lambda(\phi_1^2 + \phi_2^2)^2. \tag{2.12}$$

There are two cases to be examined here. The first where $\lambda < 0$ and $\mu^2 > 0$, shown in Figure 2.1(a), corresponds to a conventional mass term for a scalar particle, as in the Klein-Gordon Lagrangian, with a self-interaction and a unique minimum of the potential when $\phi = 0$ leaving the symmetry unbroken. However if we consider the case of both $\lambda < 0$ and $\mu^2 < 0$, shown in Figure 2.1(b), there is then a local minimum at $\phi = 0$ and now a set of true degenerate minima, which occur in the $\phi_1 - \phi_2$ plane for

$$\phi^\dagger \phi|_{\text{min}} = \frac{1}{2}(\phi_1^2 + \phi_2^2)|_{\text{min}} = \frac{\mu^2}{\lambda} \equiv v^2. \tag{2.13}$$

This gives a wrong-sign (imaginary) mass term, corresponding instead to an interaction [16], and a non-vanishing vacuum expectation value (vev) of the complex field

$$\langle 0|\phi|0\rangle = \frac{v}{\sqrt{2}}. \tag{2.14}$$



(a) (b)

Figure 2.1: Illustrations of the scalar potential $V(\phi) = \frac{1}{2}\mu^2\phi^2 - \frac{1}{4}\lambda\phi^4$ for (a) $\lambda < 0$, $\mu > 0$ and (b) $\lambda < 0$, $\mu < 0$, where v is the vacuum expectation value.

For simplicity and without loss of generality, it is chosen that $\langle 0|\phi_1|0\rangle = v$ and $\langle 0|\phi_2|0\rangle = 0$

so the field ϕ can be rewritten as a translation to this minimum energy position as

$$\phi(x) = \frac{1}{\sqrt{2}}(v + \eta(x) + i\xi(x)). \tag{2.15}$$

Re-expressing the potential in terms of η and ξ gives

$$\begin{aligned} V(\phi) &= \frac{\mu^2}{2}[v^2 + 2v\eta + \eta^2 + \xi^2] - \frac{\lambda}{4}[v^2 + 2v\eta + \eta^2 + \xi^2]^2 \\ &= \frac{\mu^2}{2}[v^2 + 2v\eta + \eta^2 + \xi^2] - \frac{\mu^2}{4v^2}[6v^2\eta^2 + v^2\xi^2 + \dots] \\ &= -\mu^2\eta^2 + \dots \end{aligned} \tag{2.16}$$

with the dots representing constant and higher order terms in the expansion, which are not considered. Substituting this back into the Lagrangian gives

$$\mathscr{L} = \frac{1}{2}\partial_\mu\xi\partial^\mu\xi + \frac{1}{2}\partial_\mu\eta\partial^\mu\eta + \mu^2\eta^2 + \dots \tag{2.17}$$

which has generated a correct sign mass term for the scalar boson η with $m_\eta = \sqrt{-2\mu^2}$ corresponding to field direction when a non-zero vev is present. There still remains a scalar boson, ξ,with no corresponding mass term, representing the remaining degree of freedom in the direction around the circle of degenerate minima, an example of the Goldstone theorem [29].

2.1.4.2 Spontaneous Symmetry Breaking of a Local Gauge Symmetry

The next step is to consider the spontaneous symmetry breaking of a local gauge symmetry, in particular the simplest case of a $U(1)$ gauge symmetry described by

$$\mathscr{L} = (\partial^\mu - ieA^\mu)\phi^\dagger(\partial_\mu - ieA_\mu)\phi - \mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2 - \frac{1}{4}F^{\mu\nu}F_{\mu\nu}. \tag{2.18}$$

Here the regular derivative ∂_μ is replaced by the covariant derivative $D_\mu = \partial_\mu - ieA_\mu$, such that the Lagrangian is invariant under the local gauge transformation $\phi \to e^{i\chi(x)}\phi$ where the gauge field transforms as $A_\mu \to A_\mu + \frac{1}{e}\partial_\mu\alpha$. Following the symmetry breaking procedure as before, choosing $\lambda < 0$, $\mu < 0$, giving ϕ_1 a non-zero vev and translating to the

minimum energy position, gives the Lagrangian as

$$\begin{aligned}\mathscr{L} &= \frac{1}{2}(\partial^\mu - ieA^\mu)(v+\eta-i\xi)(\partial_\mu - ieA_\mu)(v+\eta+i\xi) \\ &\quad - \mu^2|v+\eta+i\xi|^2 + \lambda\left(|v+\eta+i\xi)|^2\right)^2 - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} \\ &= \frac{1}{2}\partial_\mu\xi\partial^\mu\xi + \frac{1}{2}\partial_\mu\eta\partial^\mu\eta - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} + \frac{1}{2}e^2v^2A^\mu A_\mu \\ &\quad + evA^\mu\partial_\mu\xi + \mu^2\eta^2 + \dots \end{aligned} \tag{2.19}$$

This Lagrangian now contains a massless scalar ξ, a massive scalar η with $m_\eta = \sqrt{-2\mu^2}$ and most importantly a massive vector A_μ with $m_A = ev$. An issue that arises here is the interpretation of the term $evA^\mu\partial_\mu\xi$ and the apparent increase in the number of degrees of freedom, as the now massive vector boson has gained a longitudinal polarisation state. It can be shown, via performing a local gauge transformation, that the scalar field ξ can be gauged away and does not correspond to a physical particle but rather is absorbed to provide the extra longitudinal degree of freedom. This process also removes the unwanted cross-term.

2.1.4.3 The Brout-Englert-Higgs Mechanism

Finally it is possible to consider spontaneous symmetry breaking for a $SU(2)_L \times U(1)_Y$ symmetry (as required for the SM) where the covariant derivative is defined as

$$D_\mu = \partial_\mu + \frac{i}{2}g\vec{\tau}\cdot\vec{W}_\mu + ig'\frac{Y}{2}B_\mu, \tag{2.20}$$

and a $SU(2)_L$ doublet of complex scalar Higgs fields is introduced,

$$\Phi = \begin{pmatrix}\phi^+ \\ \phi^0\end{pmatrix}, \qquad \phi^+ = \frac{\phi_1 + i\phi_2}{\sqrt{2}}, \quad \phi^0 = \frac{\phi_3 + i\phi^4}{\sqrt{2}}. \tag{2.21}$$

The Lagrangian for the scalar fields is required to have a general renormalisable potential invariant under the $SU(2)_L \times U(1)_Y$ gauge transformation

$$\Phi \to \exp\left[-ig\frac{\tau_a}{2}\cdot\theta^a(x) - i\frac{g'}{2}\chi(x)\right]\Phi, \tag{2.22}$$

where $\vec{T}$ and Y are the local $SU(2)_L$ and $U(1)_Y$ gauge transformations respectively and $\vec{\tau}$ are the $SU(2)$ generators (the Pauli matrices). This gives

$$\mathscr{L}_\Phi = (D_\mu\Phi)^\dagger D^\mu\Phi + \mu^2\Phi^\dagger\Phi - \lambda(\Phi^\dagger\Phi)^2. \tag{2.23}$$

As before $\lambda < 0$ and $\mu^2 < 0$ are chosen such that a wrong-sign mass term will appear in the Lagrangian, producing a set of $SU(2)_L \times U(1)_Y$ invariant minima below $V(\Phi) = 0$ specified by

$$\left.\frac{dV}{d(\Phi^\dagger\Phi)}\right|_{min} = 0 = \mu^2 - 2\lambda(\Phi^\dagger\Phi), \tag{2.24}$$

giving

$$\Phi^\dagger\Phi|_{min} = \frac{1}{2}(\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2) = \frac{\mu^2}{2\lambda}. \tag{2.25}$$

A vacuum can then be picked from the set of minima such that $\langle 0|\phi_1|0\rangle = \langle 0|\phi_2|0\rangle = \langle 0|\phi_3|0\rangle = 0$ and a non-zero vev is assigned to ϕ_3,

$$\langle 0|\phi_3|0\rangle^2 = v^2 = \frac{\mu^2}{\lambda}. \tag{2.26}$$

In principle the choice of the vacuum direction is arbitrary, however it will be seen later that in order for the photon to remain massless the vev should be assigned to a neutral field. Expanding around this vacuum by eliminating the fields with zero vevs and setting $\phi_3 = h(x) + v$, where $h(x)$ is the Higgs field, Φ is now expressed as

$$\Phi = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ h(x) + v \end{pmatrix}. \tag{2.27}$$

The masses of the gauge bosons can now be spontaneously generated by substituting the above into the Lagrangian. From Eq. (2.20) this gives

$$\begin{aligned} D_\mu\Phi &= \frac{1}{\sqrt{2}}\begin{pmatrix} \partial_\mu + i\frac{g}{2}W_\mu^3 + i\frac{g'}{2}B_\mu & i\frac{g}{2}(W_\mu^1 - iW_\mu^2) \\ i\frac{g}{2}(W_\mu^1 + iW_\mu^2) & \partial_\mu + i\frac{g}{2}W_\mu^3 + i\frac{g'}{2}B_\mu \end{pmatrix}\begin{pmatrix} 0 \\ h + v \end{pmatrix} \\ &= \frac{1}{\sqrt{2}}\begin{pmatrix} i\frac{g}{2}(W_\mu^1 - iW_\mu^2)(h + v) \\ (\partial_\mu - i\frac{g}{2}W_\mu^3 + i\frac{g'}{2}B_\mu)(h + v) \end{pmatrix}. \end{aligned} \tag{2.28}$$

Using the relations associating the fields W_μ^i and B_μ with physical fields

$$\begin{aligned} W_\mu^+ &= \frac{1}{\sqrt{2}}(W_\mu^1 + iW_\mu^2) \\ W_\mu^- &= \frac{1}{\sqrt{2}}(W_\mu^1 - iW_\mu^2) \\ Z_\mu &= \cos\theta_W W_\mu^3 - \sin\theta_W B_\mu \\ A_\mu &= \sin\theta_W W_\mu^3 + \cos\theta_W B_\mu \end{aligned} \tag{2.29}$$

and $\cos\theta_W = g^2/\sqrt{(g^2+g'^2)}$, the covariant derivate acting on Φ can be rewritten as

$$D_\mu\Phi = \frac{1}{\sqrt{2}}\begin{pmatrix} \frac{ig}{\sqrt{2}}W_\mu^-(h+v) \\ \left(\partial_\mu - \frac{i}{2}(g^2+g'^2)^{\frac{1}{2}}Z_\mu\right)(h+v) \end{pmatrix}. \tag{2.30}$$

A key consequence of this is the absence of any terms involving the photon field A_μ and therefore lack of photon mass term in the Lagrangian, guaranteeing the photon remains massless as the $SU(2)_L \times U(1)_Y$ symmetry has been spontaneously broken with a residual $U(1)_{em}$ symmetry. The conjugate of the covariant derivative is given by

$$(D_\mu\Phi)^\dagger = \frac{1}{\sqrt{2}}\begin{pmatrix} -\frac{ig}{\sqrt{2}}W_\mu^+(h+v) & \left(\partial_\mu + \frac{i}{2}(g^2+g'^2)^{\frac{1}{2}}Z_\mu\right)(h+v) \end{pmatrix} \tag{2.31}$$

and the Lagrangian in the unitary gauge with the chosen vev is then

$$\begin{aligned} \mathscr{L}_\Phi &= (D_\mu\Phi)^\dagger D^\mu\Phi + \mu^2\Phi^\dagger\Phi - \lambda(\Phi^\dagger\Phi)^2 \\ &= \frac{1}{2}\partial_\mu h\partial^\mu h + \frac{1}{4}g^2(h^2+2vh+v^2)W_\mu^+W^{-\mu} \\ &\quad + \frac{1}{8}(g^2+g'^2)(h^2+2vh+v^2)Z_\mu Z^\mu \\ &\quad - \mu^2h^2 - \frac{\lambda}{4}(h^4+4vh^3-v^4). \end{aligned} \tag{2.32}$$

As before a massive Higgs scalar is present, h, with $m_h = \sqrt{-2\mu^2}$ and as required it is now possible to find masses for the $W^\pm$ and Z bosons, by identifying terms of the type $M_W^2W_\mu^+W^{-\mu}$ and $\frac{1}{2}M_Z^2Z_\mu Z^\mu$, as

$$M_W = \frac{1}{2}gv, \tag{2.33}$$

$$M_Z = \frac{1}{2}(g^2+g'^2)^{\frac{1}{2}}v = \frac{1}{2}\frac{gv}{\cos\theta_W}. \tag{2.34}$$

The Lagrangian also contains interaction terms relating to VVh and $VVhh$ interactions as well as tri- and quad-linear Higgs self-interactions, h^3 and h^4, the former of which allows for non-resonant di-Higgs production discussed in Sec. 2.2. Another important consequence is by measuring the vector boson masses, M_W and M_Z, and the fine structure constant, $\alpha = e^2/4\pi$, it is possible to determine θ_W, g and v but not μ and as such the Higgs mass is not directly constrained by other experimentally measured parameters.

2.1.5 Yukawa Mass Terms

The final pieces missing from the SM are the quark and lepton masses. This is achieved through the inclusion of Yukawa interactions [16] between the fermion fields and the Higgs

fields, which have the form

$$\mathscr{L}_Y = \sum_{l=e,\mu,\tau} -G_l \left[\overline{\psi}^{\mathrm{l}}_L \Phi \psi^{\mathrm{l}}_R + \overline{\psi}^{\mathrm{l}}_L \Phi^\dagger \psi^{\mathrm{l}}_R\right] + \sum_{f=1,2,3} - \left[G_u \overline{\psi}^f_L \Phi^c \psi^f_R + G_d \overline{\psi}^f_L \Phi \psi^f_R\right] \tag{2.35}$$

where G_l are the lepton Yukawa couplings and G_u and G_d are the up and down type quark Yukawa couplings, giving the strength of the Higgs-fermion interactions. To generate the masses of the upper components of the chiral quark doublets the conjugate scalar field is required, defined as

$$\Phi^c = i\tau_2 \Phi^* = \begin{pmatrix} \phi^{0\dagger} \\ -\phi^- \end{pmatrix} = \begin{pmatrix} h+v \\ 0 \end{pmatrix}\Bigg|_{\text{unitary gauge}} . \tag{2.36}$$

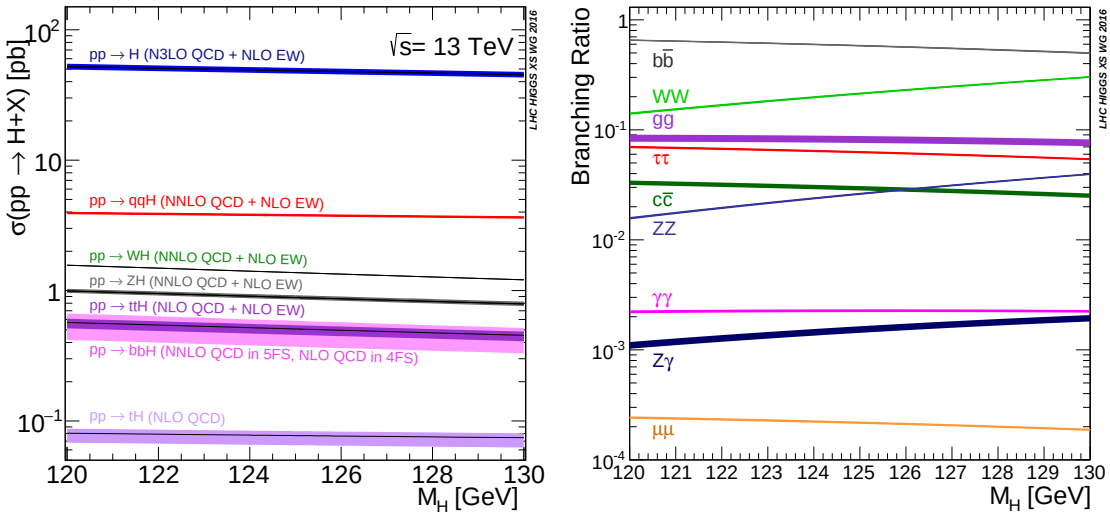
Expanding this in the unitary gauge, as in Eq. (2.36), gives terms of the form

$$-\frac{G_i}{\sqrt{2}}(\overline{\psi}_L \psi_R + \overline{\psi}_R \psi_L)(h+v) = -\frac{G_i v}{\sqrt{2}}\overline{\psi}\psi - \frac{G_i}{\sqrt{2}}\overline{\psi}\psi h, \tag{2.37}$$

where the first are identified as the fermion mass terms, $m_i = G_i v/\sqrt{2}$, and the second are the fermion-Higgs interactions. An important thing to note here is that for the unitary gauge the upper component of the lepton doublet does not couple to the Higgs field and as such does not obtain a mass, meaning the neutrinos remain massless as required by the SM.

2.2 Higgs Production

For a centre of mass energy $\sqrt{s} = 13\,\text{TeV}$ at the LHC there are four main production mechanisms for the Higgs boson: gluon-gluon fusion, vector boson fusion, associated production with a vector boson and associated production with $t\bar{t}(b\bar{b})$ quark pair. The latest theoretical results for the cross-sections of these production mechanisms as a function of the Higgs mass are shown in Figure 2.2a [6].



(a) (b)

Figure 2.2: (a) Production cross-sections and (b) decay branching ratios for the Standard Model Higgs boson as a function of the Higgs mass. [6]

2.2.1 Gluon-Gluon Fusion

The overwhelmingly dominant production process for the Higgs boson, at least one order of magnitude higher than the others, is gluon-gluon fusion (ggF), Figure 2.3a. This process is mediated by a virtual top quark loop as it has a large coupling to the Higgs boson, though other quarks contribute in the loop but are suppressed proportional to m_q^2.

2.2.2 Vector Boson Fusion

The second most frequent process, though it has a cross-section an order of magnitude smaller than the ggF process, is vector boson fusion (VBF). This occurs via a $t-$ or $u-$channel scattering of two (anti-)quarks exchanging $W^{\pm}$ or Z^0 bosons which fuse to produce a Higgs boson, shown in Figure 2.3b. In the detector this process has a distinctive signature of two forward high energy jets from the scattered quarks and as the weak bosons are colour singlets, gluon radiation in the central region of the detector is also suppressed.

These features give a good handle with which to differentiate between the VBF signal and large QCD backgrounds, making this a relatively clean channel.

2.2.3 Vector Boson Associated Production

The next highest cross-section is vector boson associated Higgs production, referred to as VH, where an incoming quark anti-quark pair produce a $W^{\pm}/Z^0$ boson which subsequently radiates a Higgs. The leading order Feynman diagram is shown in Fig. 2.3c. The decay of the $W^{\pm}/Z^0$ boson can be used to characterise these events, in particular for those events where the Higgs decays to a $b\bar{b}$ pair to reduce the QCD backgrounds, again providing a relatively clean channel to study the Higgs.

2.2.4 $t\bar{t}(b\bar{b})$ Associated Production

The final relevant single Higgs production mechanism at the LHC is a Higgs produced in association with a $t\bar{t}$ or $b\bar{b}$ pair, Fig. 2.3d. In this process $t\bar{t}/b\bar{b}$ quarks are pair produced from gluons with one pair fusing to produce a Higgs leaving two quarks in the final state. At $\sqrt{s} = 13\,\text{TeV}$ the $t\bar{t}$ and $b\bar{b}$ have similar cross-sections as although the top quark has a strong coupling to the Higgs there is a large energy requirement to produce two top anti-top pairs.

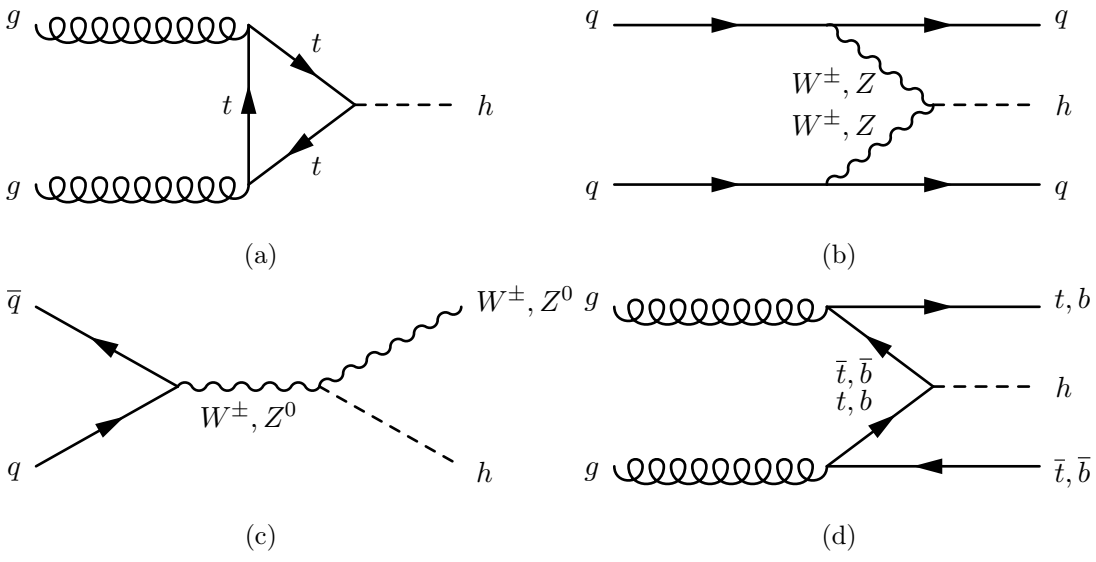


Figure 2.3: Leading order Feynman diagrams for Higgs boson production at the LHC by the (a) gluon-gluon fusion, (b) vector boson fusion, (c) vector boson associated production and (d) $t\bar{t}(b\bar{b})$ quark pair associated processes.

2.3 Di-Higgs Production

Within the SM, related to the vh^3 terms in Eq. (2.32), it is possible for the pair production of two Higgs bosons. This will provide an important test of electroweak symmetry breaking as it gives access to the trilinear Higgs self coupling, $\lambda_{hhh} = m_h^2/2v^2$, allowing for a more detailed determination of the shape of the Higgs potential. The dominant production mechanism is via the gluon-gluon fusion process with contributions from two diagrams at leading order; the top mediated triangle, Fig. 2.4a, which is sensitive to λ_{hhh} and a top mediated box, Fig. 2.4. These two diagrams destructively interfere lowering the overall cross-section dramatically with the current theory calculation to next-to-next-to-log-leading order (NNLL) for a 125.09 GeV SM Higgs boson at $\sqrt{s} = 13$ TeV given by $\sigma_{\text{NNLL}} = 33.49^{+1.44}_{-2.00}(\text{scale}) \pm 1.67(\text{theory}) \pm 0.77(\alpha_s) \pm 0.70(\text{PDF})$ fb [6].

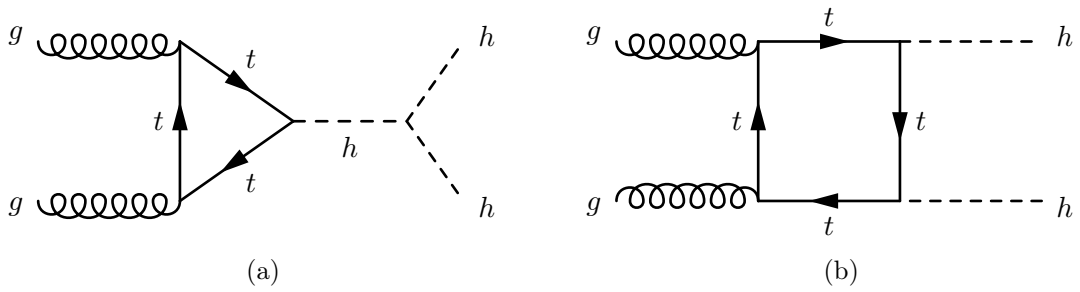


Figure 2.4: Leading order Feynman diagrams contributing to Higgs boson pair production.

Although the cross-section in the SM is very small, and there will not be enough data to achieve discovery sensitive until the High Luminosity LHC (HL-LHC), there are several possible extensions to the SM which could lead to this being enhanced. The first such BSM extension is through modifications to the trilinear Higgs couplings or to the top Yukawa coupling, changing the amount of interference between the two leading order diagrams, potentially increasing or decreasing the cross-section of non-resonant production. Higher order corrections to the potential could also lead to modifications in λ_{hhh} that could alter the cross-section prediction.

The other type of theory is the production of heavy resonances that could decay to a pair of SM $m_h = 125$ GeV Higgs bosons. One such model is that of Kaluza-Klein (KK) gravitons, G^*_{KK}, in the bulk Randall-Sundrum (RS) model [8][7]. This predicts an additional spin-2 particle, the graviton, that exists in a 5-dimensional warped geometry spacetime with constant negative curvature, with all other SM particles confined to a (3+1)-dimensional spacetime and the extra KK spatial dimension is very small and circular. The graviton masses and couplings to the SM are predicted to be at the TeV scale, with the decays of the graviton dominated by the top quark and Higgs. The partial decay width

to a pair of Higgs bosons is proportional to

$$\Gamma(G^*_{KK} \to hh) \propto \frac{k^2}{M^2_{\mathrm{Pl}}} m_G, \tag{2.38}$$

where k is the curvature scale of the warped spacetime and M_{Pl} is the effective four-dimensional Planck scale.

Another set of possible models producing a resonance which could decay to a pair of Higgs bosons are the two-Higgs-doublet models (2HDM) [9][10]. This involves the extension of the SM to include an additional Higgs doublet, one of the most minimal extensions, which contributes five physical particles. The first of these are the charge parity (CP)-even neutral Higgs bosons h and H, where by convention the former is lighter than the latter and is associated with the Higgs boson observed at the LHC. There is also a CP-odd (psuedoscalar) neutral particle A and a pair of charged Higgs $H^{\pm}$. These models have a total of six free parameters: four Higgs masses (m_h, m_H, m_A, $m_{H^{\pm}}$), the ratio of the vacuum expectation values of the two doublets ($\tan\beta$) and a Higgs mixing angle (α). Different types of the 2HDM exist where the Higgs bosons couple to a subset of the SM fermions and bosons depending on the values of α and β, with a special case occurring when $\cos(\beta-\alpha) = 0$ meaning the light CP-even h has the couplings of the SM Higgs boson. In the theory the presence of a Hhh vertex allows for the decay of the heavier CP-even Higgs to a pair of the lighter CP-even Higgs.

2.3.1 Di-Higgs to $b\,b\,\tau_{\mathrm{lep}}\,\tau_{\mathrm{had}}$

The Higgs boson has an extremely short lifetime decaying almost immediately into a number of possible fermionic and bosonic final states, including massless gluons and photons via loop processes. The branching ratios (BRs) of the possible decay modes are fixed by the mass of the Higgs itself and are defined by the ratio of the partial decay width, Γ, for a given decay process, X_i, and the total decay width as

$$BR\,(h \to X_i) = \frac{\Gamma\,(h \to X_i)}{\sum_j \Gamma\,(h \to X_j)}. \tag{2.39}$$

The SM Higgs boson decay BRs for a number of processes as a function of Higgs mass are shown in Figure 2.2b.

For the production of a Higgs boson pair there are many combinations of final state particles from the Higgs decays, shown with the corresponding branching fractions in Figure 2.5. One of the most promising decay modes in the search for di-Higgs processes is the $b\bar{b}\tau^+\tau^-$ final state. This has the third highest branching fraction at 7.3%, however

compared to the channels with higher branching fractions, $b\bar{b}b\bar{b}$ and $b\bar{b}W^+W^-$, it provides a relatively clean signature owing to better separation between the dominant $t\bar{t}$ and multi-jet backgrounds.

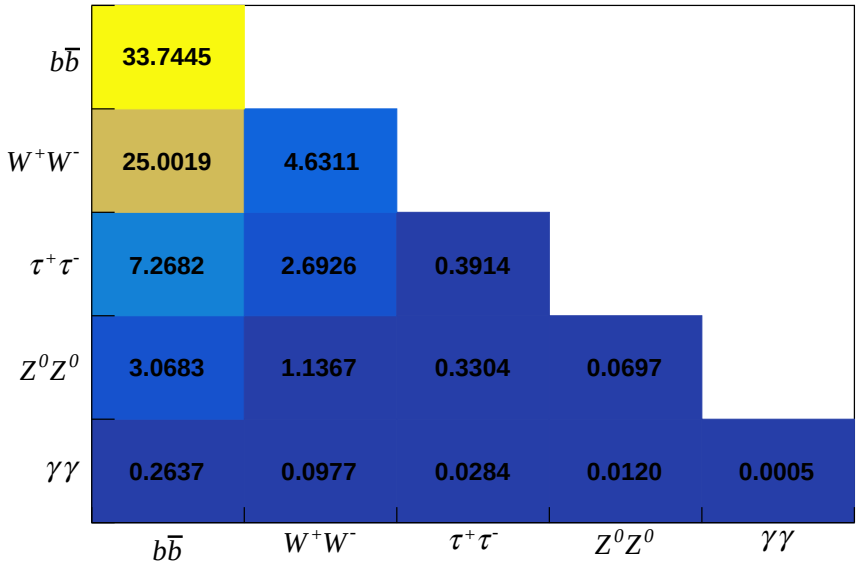


Figure 2.5: Final state branching fractions (%) for the decays of pair produced Higgs bosons.

The taus can decay to either leptons or hadrons via the weak interaction. The hadronic decays are categorised as either 1- or 3-prong depending on the number of charged pions present in the decay, with the total branching fraction for hadronic tau decays being 64.79% of all tau decays. The branching fractions for leptonic decays to a muon or electron and neutrinos are 17.39% and 17.82% respectively. The dominant decay channels of taus are summarised in Table 2.3. The combination of the electron and muon channels $\tau_{\text{lep}}\tau_{\text{had}}$ makes up almost half of the Higgs to di-tau decays, with $BR(h \to \tau_e\tau_{\text{had}}) = 0.2309$ and $BR(h \to \tau_\mu\tau_{\text{had}}) = 0.2253$, and will be the focus of the analysis in this thesis. The nomenclature τ_ℓ or τ_{lep} are used to refer to the leptonically decaying tau and τ_{had} refers to the hadronically decaying tau throughout.

Combining the cross-section and branching ratios, for an integrated luminosity of 36.1 fb^{-1}, the total expected number of events for standard model di-Higgs production is ~ 1200 with approximately 40 events in the $bb\tau_{\text{lep}}\tau_{\text{had}}$ final state.

Tau Decay	Branching Ratio
Hadronic 1-prong	
$\pi^{\pm}\,\nu_{\tau}$	0.1083
$\pi^{\pm}\,\pi^{0}\,\nu_{\tau}$	0.2552
$\pi^{\pm}\,\pi^{0}\,\pi^{0}\,\nu_{\tau}$	0.0930
$\pi^{\pm}\,\pi^{0}\,\pi^{0}\,\pi^{0}\,\nu_{\tau}$	0.0105
Hadronic 3-prong	
$\pi^{\pm}\,\pi^{\pm}\,\pi^{\mp}\,\nu_{\tau}$	0.0899
$\pi^{\pm}\,\pi^{\pm}\,\pi^{\mp}\,\pi^{0}\,\nu_{\tau}$	0.0270
Total Hadronic	0.6479
Leptonic	
$e\,\overline{\nu}_{e}\,\nu_{\tau}$	0.1782
$\mu\,\overline{\nu}_{\mu}\,\nu_{\tau}$	0.1739

Table 2.3: Branching ratios of the dominant decay channels of taus. 1- and 3-prong refer to the number of charged pions present in the tau decay.[19]

2.4 Important Background Processes

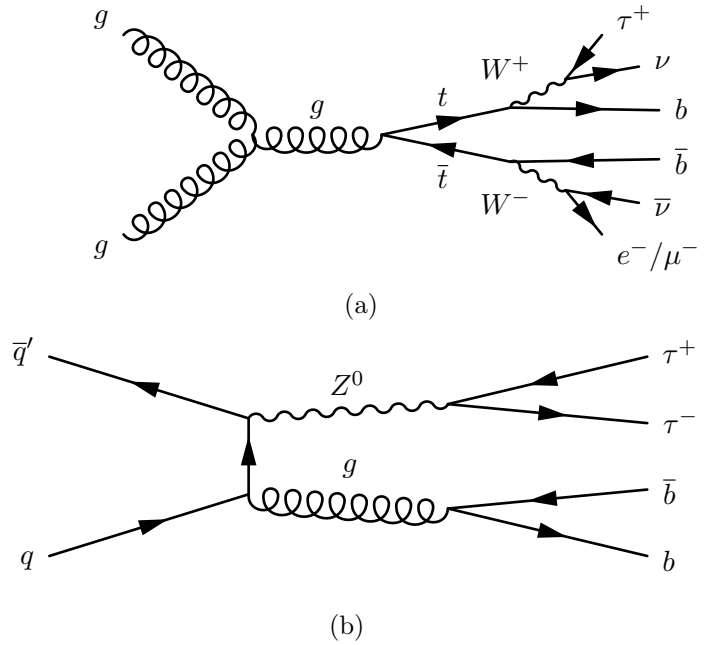


Figure 2.6: Examples of leading order Feynman diagrams for the (a) $t\bar{t}$ and (b) $Z^0 \to \tau^+\tau^- + b\bar{b}$ background processes.

The dominant SM background to the search for di-Higgs to $bb\tau\tau$, owing to its large production cross-section at the LHC (predicted to be ~ 834 pb at $\sqrt{s} = 13\,\text{TeV}$ [30]) and potentially similar final state, is the production of top-antitop ($t\bar{t}$) quark pairs. Produced predominantly through gluon-gluon fusion at the LHC, shown in Figure 2.6(a), the top quark decays almost exclusively to a b-quark and a W boson, with the final state then being characterised by the decay of the W boson. For the analysis these are categorised into two types of events; those with a real hadronically decaying tau from one of the W decays, and those where one of the W's decays hadronically and one of the produced jets is misidentified as a hadronically decaying tau. In both cases the other W boson is required to decay to a light lepton. The modelling and estimation of these backgrounds will be discussed in later chapters.

Another important background is SM production of a Z^0 boson in association with b-quark initiated jets [31], where the Z^0 boson decays to a pair of taus ($Z^0 \to \tau^+\tau^- + b\bar{b}$) showin in Figure 2.6(b). This presents a more challenging background to distinguish from the signal than the $t\bar{t}$ as it has an almost identical final state to the signal processes, with each pair of final state particles being produced together. There are however some handles which can be used to characterise this process, such as the di-tau system mass should be reconstructed to the Z^0 mass instead of the Higgs mass, which a multi-variate method could make use of as discussed later.

Chapter 3

Machine Learning Theory

Machine Learning (ML) methods have grown significantly in popularity in High Energy Physics (HEP) over recent years, enabling the search for small signals amongst large backgrounds. Currently Multi-Variate Analysis (MVA) techniques are most commonly used as binary classifiers to combine multiple discriminating variables into one final discriminator, which offers better separation than the individual variables by making use of the higher dimensionality and correlations between variables.
MVAs can be utilised in two ways to construct a search for a new signal process; for event selection by putting a cut on the MVA output of each event such that one side of the cut is more signal like and the other is more background like, or to construct a test statistic using the output distribution for all events to perform a hypothesis test to accept or reject the presence of the signal process. For the purpose of the analysis presented in this thesis the later is used. Due to the Neyman-Pearson lemma [32] there is an optimal classifier which leads to an upper limit on the achievable improvements by changing between MVA methods, however until this is reached it is worth exploring potential improvements.
There are many tools available that implement MVA techniques with the most popular in particle physics being TMVA [33], which is used for the analyses in this thesis. This chapter describes the foundations behind two supervised ML methods considered in this thesis, Boosted Decision Trees (Sec. 3.1) and Support Vector Machines (Sec. 3.2). Generalisation techniques (Sec. 3.3) used for training that can be applied to any classifier are also discussed. Further information and more detailed discussion on these topics can be found in References [34], [35] and [36].

3.1 Boosted Decision Trees

Boosted Decision Trees (BDTs) are probably one of the most frequently used classifiers in High Energy Physics (HEP) at this time. This is mainly due to the relative transparency and simplicity to understand this algorithm compared to other modern multi-variate techniques.

3.1.1 Decision Tree Learning

A Decision Tree (DT) is a structure of repeated binary (pass/fail) cuts made on individual input variables that continue until a termination criteria is satisfied. At each node the variable to be cut on and the value of the cut are optimised to give the best separation between the signal and background datasets using various separation criteria, such as the Gini Index, $p \cdot (1 - p)$ where p is the signal purity. This is the default in TMVA. The route down each branch to a final node (or leaf) represents a different subset of the data in hyperspace where the type of that subset depends on the purity of signal or background in the final node. This is a key advantage of DTs over a standard cuts-based analysis which only selects one region in hyperspace. However DTs are very susceptible to overtraining [1] due to statistical fluctuations in the data, as if two variables have similar separation power a small change may cause one to be selected over the other and impact the tree from this node onwards.

3.1.2 Boosting

The problem of instability in individual DTs can be reduced through the use of Random Forests (RFs), an ensemble method that combines the outputs from different trees through some majority vote process to achieve a stronger combined learner providing better separation than the individual weaker trees. There are various approaches to training RFs such a Bootstrap Aggregating (Bagging) [37], where trainings are performed on resampled subsets of the data, or Boosting, an iterative method of reweighting misclassified events before each training, where the final output is a weighted average (usually based on accuracy) of individual DTs. The most commonly used algorithm in HEP is Adaptive Boosting (AdaBoost) [38] described in the following.

Consider a training dataset of m examples $S = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)\}$ where $\mathbf{x}_i$ is the set of input variables for data point i and $y_i \in \pm 1$ is its corresponding class label. For an individual tree the result of an event is $h_j(\mathbf{x}_i) = \pm 1$ depending on whether the leaf containing it is mainly signal or background respectively, and the tree error is given by the sum over the weights, w_i, of misclassified events

$$\epsilon_j = \sum_{y_i \neq h_j(\mathbf{x}_i)} w_i^{(j)} \Big/ \sum_{i=1}^{n} w_i^{(j)}. \tag{3.1}$$

Once a tree is trained by minimising ϵ_j the corresponding score is then

[1] Overtraining or overfitting refers to the MVA method learning the training dataset too well, as it may fit to noise or random fluctuations, such that when it is applied to an independent unseen test set the performance is significantly worse.

$$\alpha_j = \beta \ln \frac{1-\epsilon_j}{\epsilon_j}, \tag{3.2}$$

where β is a free parameter representing the learning rate to be optimised for the analysis (referred to as AdaBoost Beta). Before the next tree is trained the weights of the misclassified events are updated following

$$w_i^{(j+1)} = w_i^{(j)} e^{\alpha_j}, \tag{3.3}$$

such that the norm is preserved. The final output for an event once all trees have been trained is then given by

$$h_{fin}(\mathbf{x}_i) = \frac{1}{N_{trees}} \sum_{j=1}^{N_{trees}} \alpha_j h_j(\mathbf{x}_i). \tag{3.4}$$

DTs can also be pruned to reduce overtraining, which involves training a tree to its maximum size and depth, before then removing statistically insignificant nodes from the bottom up. However as BDTs work best with weak learners, a better approach is to limit the tree depth rather than prune individual trees and as such pruning is not used in the analyses described later in this thesis.

3.2 Support Vector Machines

One promising alternative approach to BDTs, currently used infrequently in particle physics, is that of Support Vector Machines (SVMs), a supervised learning model used with associated learning algorithms. Originally developed in the 1960s, with the current standard version proposed in 1995 [39], SVMs aim to classify data points using a maximal margin hyperplane mapped from a linear classification problem in a possibly infinite dimensional hyperspace.

3.2.1 Hard Margin SVMs

First consider the simplest problem of linear classification with the SVM where the training set, $\mathbf{x}$, is linearly separable. Similar to the process used for the Fisher discriminant [40], a separating hyperplane is defined given by

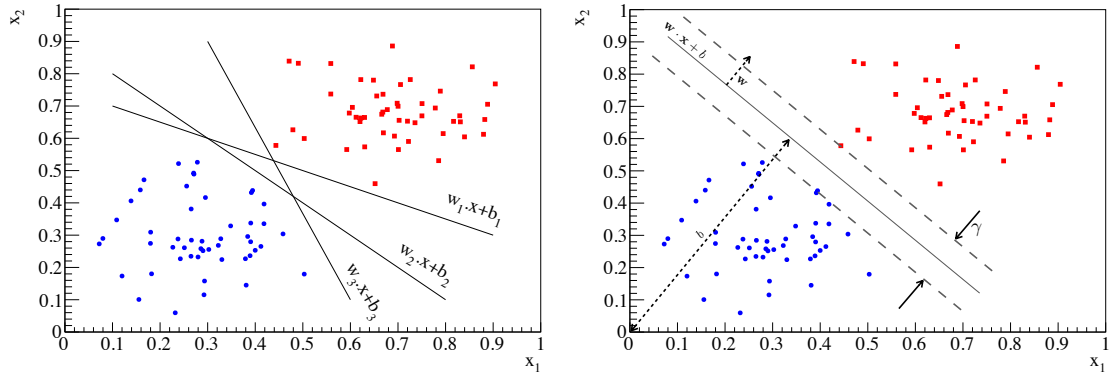
$$\langle \mathbf{w} \cdot \mathbf{x} \rangle + b = 0, \tag{3.5}$$

where $\mathbf{w}$, referred to as the weight vector, is a vector perpendicular to the hyperplane determining the orientation, and b, referred to as the bias, determines the distance of the hyperplane from the origin. The inner product is defined here as $\langle \mathbf{w} \cdot \mathbf{x} \rangle = \sum_i w_i x_i$. A decision function defined by $y_i = \text{sign}(\langle \mathbf{w} \cdot \mathbf{x_i} \rangle + b)$ is used to positively and negatively classify $\mathbf{x}_i$, the points in the training set. Without further constraint the choice of separating hyperplane is arbitrary (see Figure 3.1) and as such the functional margin, γ_i, for a data point is then defined as

$$\gamma_i = y_i(\langle \mathbf{w} \cdot \mathbf{x}_i \rangle + b) \geq 1, \tag{3.6}$$

where it is imposed that no points lie within the margins, as defined by the equality condition of the above equation.
Now considering two points which lie closest to the hyperplane, $\mathbf{x}^+$ and $\mathbf{x}^-$, they will have functional margins of plus and minus one respectively and the geometric margin γ can then be defined as

$$\begin{aligned}
\gamma &= \frac{1}{2}\left(\left\langle \frac{\mathbf{w}}{\| \mathbf{w} \|_2} \cdot \mathbf{x}^+ \right\rangle - \left\langle \frac{\mathbf{w}}{\| \mathbf{w} \|_2} \cdot \mathbf{x}^- \right\rangle\right) \\
&= \frac{1}{2 \| \mathbf{w} \|_2}\left(\langle \mathbf{w} \cdot \mathbf{x}^+ \rangle - \langle \mathbf{w} \cdot \mathbf{x}^- \rangle\right) \\
&= \frac{1}{2 \| \mathbf{w} \|_2}\left[(1-b) - (-1-b)\right] \\
&= \frac{1}{\| \mathbf{w} \|_2},
\end{aligned} \tag{3.7}$$



(a) (b)

Figure 3.1: Illustrations of separating hyperplanes for a linearly separable data set. (a) examples of the linear discriminant and arbitrary choice of hyperplane with no additional constraints, and (b) the hard margin SVM where the solid line is the separating hyperplane, described by $\mathbf{w}$ and b.

where the 2-norm, $\| \mathbf{w} \|_2 = \left(\sum_{p=1}^{q} |w_p|^2 \right)^{\frac{1}{2}}$, is a sum over the input dimensions.

For a given training set, $\mathbf{S} = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)\}$, this leads to the following optimisation problem

$$\text{minimise}_{\mathbf{w},b} \quad \langle \mathbf{w} \cdot \mathbf{w} \rangle \,, \tag{3.8}$$

$$\text{subject to} \quad y_i \left(\langle \mathbf{w} \cdot \mathbf{x}_i \rangle + b \right) \geq 1, \tag{3.9}$$

$$i = 1, \ldots, n, \tag{3.10}$$

where n is the total number of points in the dataset and without changing the solution of the problem, $\| \mathbf{w} \|_2$ has been replaced with $\langle \mathbf{w} \cdot \mathbf{w} \rangle$, as it is more difficult to solve an optimisation problem which depends on the square root of the quantity.

This optimisation problem can now be expressed in Lagrangian form [34, p.79–88], which allows the constraint, Eq. (3.9), to enter the minimisation problem, as

$$\mathcal{L}(\mathbf{w}, b, \boldsymbol{\alpha}) = \frac{1}{2} \langle \mathbf{w} \cdot \mathbf{w} \rangle - \sum_{i=1}^{n} \alpha_i \left[y_i \left(\langle \mathbf{w} \cdot \mathbf{x}_i \rangle + b \right) - 1 \right], \tag{3.11}$$

where the Lagrange multipliers $\alpha_i \geq 0$ have been introduced, as well as a factor of $1/2$ for mathematical convenience. If this primal Lagrangian is then differentiated with respect to both $\mathbf{w}$ and b, and the derivatives are set to zero, one obtains

$$\frac{\partial \mathcal{L}(\mathbf{w}, b, \boldsymbol{\alpha})}{\partial \mathbf{w}} = 0 \quad \Rightarrow \quad \mathbf{w} = \sum_{i=1}^{n} \alpha_i y_i \mathbf{x}_i, \tag{3.12}$$

$$\frac{\partial \mathcal{L}(\mathbf{w}, b, \boldsymbol{\alpha})}{\partial b} = 0 \quad \Rightarrow \quad \sum_{i=1}^{n} \alpha_i y_i = 0. \tag{3.13}$$

Substituting these relations back into Eq. (3.11), one obtains the dual Lagrangian

$$\mathcal{L}(\mathbf{w}, b, \boldsymbol{\alpha}) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j \left\langle \mathbf{x}_i \cdot \mathbf{x}_j \right\rangle, \tag{3.14}$$

which leads to the dual optimisation problem

$$\text{maximise} \quad \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j \left\langle \mathbf{x}_i \cdot \mathbf{x}_j \right\rangle, \tag{3.15}$$

$$\text{subject to} \quad \sum_{i=1}^{n} \alpha_i y_i = 0, \tag{3.16}$$

$$\alpha_i \geq 0, \quad i = 1, \ldots, n, \tag{3.17}$$

where the weight vector $\mathbf{w} = \sum_{i=1}^{n} \alpha_i y_i \mathbf{x}_i$ realises the geometric margin $\gamma = 1/ \parallel \mathbf{w} \parallel_2$. It should be noted that the geometric margin corresponds to a physical distance in the problem space, as this is an important concern when defining positive semi-definite (PSD) kernel functions, discussed in Section 3.2.3. The dual representation gives the relationship between pairs of data points, which for the case of SVM is an inner product. This key property allows for the use of Kernel functions for data sets which are not linearly separable in the input space, which is discussed in a later section.

Notice that b does not appear in the dual optimisation problem, and so is found using the primal constraints as

$$b = -\frac{\max_{y_i=-1} \left(\langle \mathbf{w} \cdot \mathbf{x}_i \rangle \right) + \min_{y_i=1} \left(\langle \mathbf{w} \cdot \mathbf{x}_i \rangle \right)}{2}. \tag{3.18}$$

Using the Karush-Kuhn-Tucker (KKT) [41][42] complementarity conditions from optimisation theory, it is stated that the optimal solution $\boldsymbol{\alpha}, (\mathbf{w}, b)$ must satisfy

$$\alpha_i [y_i \left(\langle \mathbf{w} \cdot \mathbf{x}_i \rangle + b \right) - 1] = 0, \quad i = 1, ..., l. \tag{3.19}$$

The implication is only the points $\mathbf{x}_i$ which lie closest to the hyperplane, and therefore have a functional margin of 1, correspond to non-zero α_i with all other α_i equal to zero. It is only these points which contribute to the expression of the weight vector and hence

these are known as support vectors, where their indices will be denoted as sv and l is the number of support vectors. This leads to the optimal hyperplane expressed in the dual representation becoming

$$f(\mathbf{x}, \boldsymbol{\alpha}, b) = \langle \mathbf{w} \cdot \mathbf{x} \rangle + b = \sum_{i \in \text{sv}} y_i \alpha_i \langle \mathbf{x}_i \cdot \mathbf{x} \rangle + b. \tag{3.20}$$

Again due to the KKT conditions, for $j \in \text{sv}$ the margin becomes

$$\gamma_j = y_j f(\mathbf{x}_j, \boldsymbol{\alpha}, b) = y_j \left(\sum_{i \in \text{sv}} y_i \alpha_i \langle \mathbf{x}_i \cdot \mathbf{x}_j \rangle + b \right) = 1, \tag{3.21}$$

and therefore the inner product of weight vectors becomes

$$\langle \mathbf{w} \cdot \mathbf{w} \rangle = \sum_{i,j \in \text{sv}} y_i y_j \alpha_i \alpha_j \langle \mathbf{x}_i \cdot \mathbf{x}_j \rangle \tag{3.22}$$

$$= \sum_{j \in \text{sv}} \alpha_j y_j \sum_{i \in \text{sv}} y_i \alpha_i \langle \mathbf{x}_i \cdot \mathbf{x}_j \rangle , \tag{3.23}$$

which using the definition of the margin, Eq. (3.21), gives

$$\langle \mathbf{w} \cdot \mathbf{w} \rangle = \sum_{j \in \text{sv}} \alpha_j (1 - y_j b) \tag{3.24}$$

$$= \sum_{i \in \text{sv}} \alpha_i. \tag{3.25}$$

This means that for the dual optimisation problem solved by parameters $\boldsymbol{\alpha}$ and b, the weight vector $\mathbf{w} = \sum_{i=1}^{n} y_i \alpha_i \mathbf{x}_i$ realises the maximal margin hyperplane with geometric margin

$$\gamma = \frac{1}{||\mathbf{w}||_2} = \left(\sum_{i \in \text{sv}} \alpha_i \right)^{-\frac{1}{2}}. \tag{3.26}$$

3.2.2 Soft Margin SVMs

Though the hard margin SVM, described in Section 3.2.1, is an important concept in establishing the basis for developing SVMs, this is not applicable for most real world cases. Issues arise if the data are not linearly separable due to being noisy[2] and the training can become susceptible to overfitting due to a few points in the training set, so some classification error must be allowed.

[2]Noisy refers to data having small variations between different independent samples and can lead to overfitting as the method could learn a particular feature in one dataset which means it then misclassifies points when using the other.

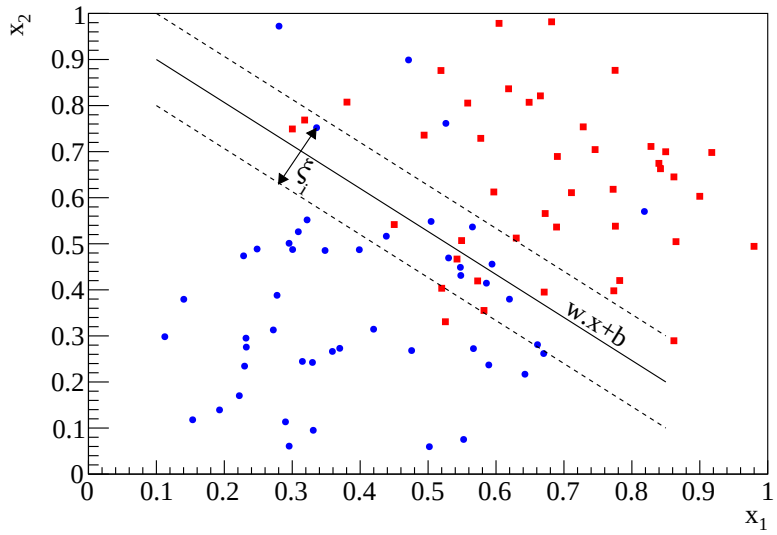


Figure 3.2: Illustration of the soft margin SVM. The slack parameters, $\boldsymbol{\xi}_i$, allow for misclassification of points weighted by the tunable cost parameter, C. Dotted lines are the margins, with the SVM finding the maximum geometric margin, γ.

In order to accept some misclassification of points, the margin constraints must be allowed to be violated. This is achieved by the introduction of slack variables, $\boldsymbol{\xi}_i$, given by

$$y_i \left(\langle \mathbf{w} \cdot \mathbf{x}_i \rangle + b \right) \geq 1 - \boldsymbol{\xi}_i, \tag{3.27}$$

$$\boldsymbol{\xi}_i \geq 0, i = 1, \ldots, n, \tag{3.28}$$

c.f. Eq. (3.6) and is illustrated in Figure 3.2.

The corresponding primal Lagrangian for the optimisation problem from generalisation theory is then

$$\mathcal{L}(\mathbf{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \mathbf{r}) = \frac{1}{2} \langle \mathbf{w} \cdot \mathbf{w} \rangle + C \sum_{i=1}^{n} \boldsymbol{\xi}_i - \sum_{i=1}^{n} \alpha_i [y_i (\langle \mathbf{x_i} \cdot \mathbf{w} \rangle + b) - 1 + \boldsymbol{\xi}_i] - \sum_{i=1}^{n} \mathbf{r}_i \boldsymbol{\xi}_i, \tag{3.29}$$

where $\alpha_i \geq 0$ and $\mathbf{r}_i \geq 0$. The parameter C, known as the cost, is a tuneable parameter that acts as a penalty for misclassifying points and should be varied through a range of values and selected via optimisation against some figure of merit (FOM). As it is varied, the value of the norm of the weight vector changes and so when choosing a particular C, a corresponding value of $||\mathbf{w}||_2$ is selected and $\boldsymbol{\xi}$ is then minimised for that value. Again to find the dual optimisation problem, the primal Lagrangian is minimised with respect to $\mathbf{w}$, b and now also $\boldsymbol{\xi}$ giving

$$\frac{\partial \mathcal{L}(\mathbf{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \mathbf{r})}{\partial \mathbf{w}} = \mathbf{0} \quad \Rightarrow \quad \mathbf{w} = \sum_{i=1}^{n} y_i \alpha_i \mathbf{x_i}, \tag{3.30}$$

$$\frac{\partial \mathcal{L}(\mathbf{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \mathbf{r})}{\partial \boldsymbol{\xi}} = 0 \quad \Rightarrow \quad C - \alpha_i - \mathbf{r}_i = 0, \tag{3.31}$$

$$\frac{\partial \mathcal{L}(\mathbf{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \mathbf{r})}{\partial b} = 0 \quad \Rightarrow \quad \sum_{i=1}^{n} y_i \alpha_i = 0. \tag{3.32}$$

Substituting these relations into the primal Lagrangian gives the same dual objective function as for the hard margin case, Eq. (3.17),

$$\mathcal{L}(\mathbf{w}, b, \boldsymbol{\xi}, \alpha, \mathbf{r}) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} y_i y_j \alpha_i \alpha_j \left\langle \mathbf{x}_i \cdot \mathbf{x}_j \right\rangle, \tag{3.33}$$

however there is now a constraint, Eq. (3.31), as $\mathbf{r}_i \geq 0$ enforces $\alpha_i \leq C$. It is also found that $\boldsymbol{\xi}_i \neq 0$ only if $\mathbf{r}_i = 0$ and so $\alpha_i = C$. This leads to the KKT complementarity conditions

$$\alpha_i [y_i (\langle \mathbf{x}_i \cdot \mathbf{w} \rangle + b) - 1 + \boldsymbol{\xi}_i] = 0, \qquad i = 1, ..., n, \tag{3.34}$$

$$\boldsymbol{\xi}_i (\alpha_i - C) = 0, \qquad i = 1, ..., n, \tag{3.35}$$

which therefore leads to the the optimisation problem

$$\text{maximise} \quad \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j \left\langle \mathbf{x}_i \cdot \mathbf{x}_j \right\rangle, \tag{3.36}$$

$$\text{subject to} \quad \sum_{i=1}^{n} \alpha_i y_i = 0, \tag{3.37}$$

$$0 \leq \alpha_i \leq C, \quad i = 1, \ldots, n. \tag{3.38}$$

3.2.3 Kernel Methods

As mentioned previously, for SVMs the data appear within inner products in both the dual objective function and the decision, which has the useful property of allowing the use of kernel functions to transform the problem into some feature space to find an optimally separating hyperplane, where one may not have previously been available.

This is done by changing the representation of the data

$$\mathbf{x} = (x_1, ..., x_n) \mapsto \boldsymbol{\phi}(\mathbf{x}) = (\phi_1(\mathbf{x}), ..., \phi_m(\mathbf{x})), \tag{3.39}$$

which is equivalent to mapping the input space X into some new feature space $F = \{\boldsymbol{\phi}(\mathbf{x})|\mathbf{x} \in X\}$, where m is not necessarily equal to n as the feature space may have different dimensionality to the input space. A kernel function is defined as a function K, such that for all $\mathbf{x}, \mathbf{z} \in X$

$$K(\mathbf{x}, \mathbf{z}) = \langle \boldsymbol{\phi}(\mathbf{x}) \cdot \boldsymbol{\phi}(\mathbf{z}) \rangle \,, \tag{3.40}$$

from which it is clear that the inner product is required to allow for the kernel to be used, with $K(\mathbf{x}, \mathbf{z}) = \langle \mathbf{x} \cdot \mathbf{z} \rangle$ a kernel function where the feature map is the the identity. A key feature of kernel functions is that the feature map or feature space do not need to be known, as you are only required to calculate the inner product of these spaces. This leads to some required properties to define a proper kernel function such that it corresponds to some feature space. The kernel function should clearly be symmetric

$$K(\mathbf{x}, \mathbf{z}) = \langle \boldsymbol{\phi}(\mathbf{x}) \cdot \boldsymbol{\phi}(\mathbf{z}) \rangle = \langle \boldsymbol{\phi}(\mathbf{z}) \cdot \boldsymbol{\phi}(\mathbf{x}) \rangle = K(\mathbf{z}, \mathbf{x}), \tag{3.41}$$

and should also satisfy the Cauchy-Schwartz inequality

$$\begin{aligned} K(\mathbf{x}, \mathbf{z})^2 = \langle \boldsymbol{\phi}(\mathbf{x}) \cdot \boldsymbol{\phi}(\mathbf{z}) \rangle^2 \leq & \| \boldsymbol{\phi}(\mathbf{x}) \|^2 \| \boldsymbol{\phi}(\mathbf{z}) \|^2 & (3.42) \\ = \langle \boldsymbol{\phi}(\mathbf{x}) \cdot \boldsymbol{\phi}(\mathbf{x}) \rangle \langle \boldsymbol{\phi}(\mathbf{z}) \cdot \boldsymbol{\phi}(\mathbf{z}) \rangle = & K(\mathbf{x}, \mathbf{x}) K(\mathbf{z}, \mathbf{z}). & (3.43) \end{aligned}$$

These alone however are not strong enough conditions to guarantee the existence of a feature space. This is guaranteed by a kernel meeting Mercer's condition [43] which ensures that the kernel will be PSD, hence giving real valued outputs. These properties allow for kernel functions to be defined directly, implicitly defining the feature space, rather than having to define a complicated feature space and calculate the kernel function from that. It follows that kernels can also be defined from other kernels, for example if we consider two symmetric PSD kernel functions $K_1(\mathbf{x}, \mathbf{z})$ and $K_2(\mathbf{x}, \mathbf{z})$, it is obvious that $K(\mathbf{x}, \mathbf{z}) = K_1(\mathbf{x}, \mathbf{z}) + K_2(\mathbf{x}, \mathbf{z})$ and $K(\mathbf{x}, \mathbf{z}) = K_1(\mathbf{x}, \mathbf{z}) K_2(\mathbf{x}, \mathbf{z})$ will also lead to new PSD symmetric kernels.

The application in SVMs is the replacement of the inner product with the kernel function, leading to the objective function, Eq. (3.36), becoming

$$\text{maximise} \quad \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j). \tag{3.44}$$

3.2.4 Sequential Minimal Optimisation

In order to find the vector $\boldsymbol{\alpha}$, an algorithm is required for the optimisation problem. The method commonly used (and the one implemented in TMVA), is that of Sequential Minimal Optimisation (SMO) which takes the concept of decomposition methods to the extreme, only adjusting pairs of α_i in each iteration. This allows for an analytic solution to be found,

and although optimisation may require many iterations, only simple calculations need to be performed at each step which allows the optimisation to be reasonably quick.
The following discusses firstly how the optimisation is performed analytically and then how the choice of the two α's is determined by a pair of heuristics.

3.2.4.1 SMO - Finding the Analytical Solution

Firstly its is clear that the condition from Eq. (3.32) must be enforced, such that when one Lagrange multiplier is updated, another must also be updated to maintain the condition. Consider that the Lagrange multipliers α_i and α_j have been chosen, which are bounded by the linear constraint as

$$\alpha_i^{new} y_i + \alpha_j^{new} y_j = \alpha_i^{old} y_i + \alpha_j^{old} y_j = \text{constant}. \tag{3.45}$$

They are also bound within the box defined by $0 \leq \alpha_i, \alpha_j \leq C$, which is a constraint from the dual optimisation problem. These constraints place restrictions on how the multipliers can then be updated, so if α_j^{new} is computed first it must satisfy

$$U \leq \alpha_j^{new} \leq V, \tag{3.46}$$

where

$$U = \begin{cases} \max(0, \alpha_j^{old} - \alpha_i^{old}) & \text{for } y_i \neq y_j \\ \max(0, \alpha_j^{old} + \alpha_i^{old} - C) & \text{for } y_i = y_j \end{cases}, \tag{3.47}$$

$$V = \begin{cases} \min(C, C - \alpha_j^{old} + \alpha_i^{old}) & \text{for } y_i \neq y_j \\ \min(C, \alpha_j^{old} + \alpha_i^{old}) & \text{for } y_i = y_j \end{cases}. \tag{3.48}$$

The maximum of the objective function is then achieved (the details and proof can be found in [34, p.137–140]) by computing

$$\alpha_j' = \alpha_j^{old} + \frac{y_j(E_i - E_j)}{\kappa}, \tag{3.49}$$

where E_i is the difference between the output of the function for the current hypothesis and the target classification type, defined by

$$E_i = \left(\sum_{k=1}^{n} \alpha_k y_k K(\mathbf{x}_k, \mathbf{x}_i) + b \right) - y_i, \tag{3.50}$$

and κ is the second derivative of the objective function along the diagonal line, expressed as

$$\kappa = \| \phi(\mathbf{x}_i) - \phi(\mathbf{x}_j) \|^2 = K(\mathbf{x}_i, \mathbf{x}_i) + K(\mathbf{x}_j, \mathbf{x}_j) - 2K(\mathbf{x}_i, \mathbf{x}_j). \tag{3.51}$$

The value for α_j^{new} is then selected as

$$\alpha_j^{new} = \begin{cases} V, & \text{if } \alpha_j' > V \\ \alpha_j', & \text{if } U \leq \alpha_j' \leq V \\ U, & \text{if } \alpha_j' < U \end{cases}, \tag{3.52}$$

and then it follow from the constraints that

$$\alpha_i^{new} = \alpha_i^{old} + y_i y_j (\alpha_j^{old} - \alpha_j^{new}). \tag{3.53}$$

3.2.4.2 SMO - Selection Heuristic

To try to maximise the speed of convergence, the pair of α's should be chosen in such a way that they have a large impact towards the overall progress of the optimisation. However, it is clear that this must be sufficiently simple that the computation required is less than that saved by reducing the number of iterations in order to improve the rate of convergence.
In the case of the SMO there are two selection criteria, one selecting the first point and then based on this, the second point is selected. The first point is selected by looping over the training set looking for any α_i which violate the KKT conditions, as these are the ones which require updating. When one is found a second point, α_j, must then be selected such that there is a large change to help progress the optimisation. This is done by finding another point violating the KKT conditions and which maximises $|E_1 - E_2|$, where E_i is defined in Eq. (3.50), though if one cannot be found that significantly progresses the optimisation, points from the whole training set are considered. Once a pair has been updated, the optimisation returns to looping over the KKT violating points and updating them until all points satisfy the conditions or a maximum number of iterations has been reached.

3.2.5 Implementation in TMVA

Initially the implementation of SVMs within TMVA was fairly limited, only containing the Radial Basis Function (RBF) kernel function, Eq. (3.54), which only has one tunable parameter, Γ. It was also required that hyper-parameter optimisation was done by hand, making it not particularly user friendly.
The SVM classifier is booked in TMVA with the command

```
factory->BookMethod( TMVA::Types::kSVM, "SVM", "<options>" ),
```

where it is recommended that the option `VarTransform=Norm` is set such that the input variables are scaled between ± 1, otherwise large scale variables (such as particle momentum) will dominate the training over small scale variables (for example b-tagging score) leading to pathologies in the output distributions.

The first extension was to include a KF which is the product of Gaussians with a tuneable parameter for each input variable, referred to here as the Multi-Gaussian kernel, Eq. (3.55). The reasoning for this choice is for input distributions of varying widths, where there may be a tension when trying to find the "optimum" single width parameter, while with multiple width parameters the variables can be mapped into more appropriate feature spaces.

Another common choice of kernel which has now been implemented in TMVA is the polynomial kernel, Eq. (3.56), specified by an integer order parameter, d, and an offset parameter, θ, which is required to be positive in order for the kernel to be PSD. The RBF, Eq. (3.54), Multi-Gaussian, Eq. (3.55), and Polynomial, Eq. (3.56), KFs are calculated as

$$K(\mathbf{x}, \mathbf{z}) = \exp\left(-\Gamma \parallel \mathbf{x} - \mathbf{z} \parallel^2\right), \tag{3.54}$$

$$K(\mathbf{x}, \mathbf{z}) = \prod_i \exp\left(-\Gamma_i |x_i - z_i|^2\right), \tag{3.55}$$

$$K(\mathbf{x}, \mathbf{z}) = (\mathbf{x} \cdot \mathbf{z} + \theta)^d. \tag{3.56}$$

For the RBF and Polynomial KFs the values of the hyper-parameters can be specified with the `Gamma`, and the `Theta` and `Order` options respectively. The Multi-Gaussian KF requires the `GammaList` option followed by a list of Γ's in the same order as the input variables to be specified. As discussed in Sec. 3.2.3, valid KFs can also be formed from both taking the products and sums of other KFs. These have also been implemented within TMVA to allow the product or sum of the above kernels to be calculated and used as a kernel. To use either of these the kernel should be set to `Prod` or `Sum` and then using the `MultiKernels` option, a list of kernels separated with `*/+` delimiters for the product or sum respectively should be supplied.

Following the implementation for the BDTs within TMVA, hyper-parameter optimisation has also been included for SVMs. Ideally using a statistically independent validation sample the optimisation can be done either by linearly scanning through the parameter space, which is not scalable to large parameter spaces so suited for a kernel such as the polynomial where the range and number of hyper-parameters is potentially small, or through minimisation which can be prone to finding local minima. The parameters to be optimised and the ranges and number of steps over which the optimisation is performed can be specified in the options string when booking the method, for example `Tune=Gamma[0.1;0.4]` will optimise the width parameter, Γ, in the specified range (split into 100 steps by default if using `Scan`).

Due to the inclusion of the cost function, it was found when training with data sets where one type sample is significantly larger than the other, that the smaller sample would have little impact on the training and the larger sample would be learnt too well leading to overfitting. This is avoided with the implementation of a weighted cost function with two cost parameters C_{large} and C_{small}, where it is assumed that the chosen cost parameter is for the smaller sample. The two cost parameters are then related by

$$C_{large} = \frac{N_{small}}{N_{large}} C_{small}. \tag{3.57}$$

3.3 Generalisation

Whenever using MV methods it is important that the trained method is robust to over-training, to the end that the performance of an unseen dataset can be accurately predicted by the model and is reproducible such that a consistent result is obtained with the addition of new data or when evaluated on systematic variations. This motivates the requirement for validation techniques for two main reason, firstly for model selection, as most MV methods have at least one free parameter which has to be chosen through some form of "optimal" selection, and secondly for performance estimation to evaluate how the chosen models perform. In both cases there are various FOMs which can be constructed to compare one training to another, however these should be selected on a problem specific basis. A review of the techniques discussed in the following can be found in [44].

For unlimited datasets these issues are trivial as one can simply iterate through the parameters and select the model with the best performance based on the FOM. In reality our datasets are much smaller than they ideally would be and as such this is not possible. Naïvely the entire dataset could be used to select and train the model and evaluate the performance, however this in general leads to overfitting or overtraining as the method can learn fluctuations in the dataset and therefore performs worse on unseen data. This effect is more distinct in methods with a large number of hyper-parameters as they have more flexibility to fit to the data. This also gives an over optimistic estimation of performance.

3.3.1 Hold-out

The most common method currently used within HEP to promote generalisation and the simplest extension is hold-out validation. This involves splitting the dataset into independent training and testing samples. These two samples can be used to select the "optimal" training point by comparing the FOMs for the training and testing samples, for which the training sample will converge where as the testing sample will have a minimum.

This method is problematic and can lead to a misleading estimate of the FOM as it is very susceptible to the splitting of the data.

3.3.2 Cross-validation

Depending on the size of the dataset it may not be possible to reserve a large amount of data for testing so the hold-out validation method may not be viable. An extension to k-fold cross-validation (CV) may be necessary and is performed as follows:

1. Split the data into k equally sized, randomly sampled, independent datasets (referred to as folds);
2. Train the classifier with the data from $k-1$ of the folds;
3. Test the classifier using the remaining fold;
4. Repeat steps (ii) and (iii) for all remaining permutations.

This method has the advantage of using the whole dataset for both training and testing, with the FOM estimated using the average of the FOMs calculated for each test fold.
The question then arises of how many folds should be used? If a large number of folds is used this will lead to a good estimate of the average error rate (or which ever FOM is being used) as the bias of the estimator will be small. However this also leads to the variance of the estimator being large, and of course a long computational time as many folds have to be trained and tested. The converse to the above is true for a small number of folds but in reality the number of folds should be motivated by the size of the dataset available. Currently in particle physics it is most common to split the dataset into 2-folds, train on one and test on the other and then reverse their roles and repeat.
A further extension to CV is to initially split the dataset into three statistically independent samples for training, validation and testing. The training and validation sets are used to choose the classifier or model and to select the optimal tuning parameters, making use of k-fold CV. Once the model and parameters are chosen, the training and validation sets are combined to train the final classifiers via k-fold CV. The previously unused testing dataset can then be used to access the performance of the trained classifiers, where their outputs can be combined via some averaging or majority voting technique.

Chapter 4

The Large Hadron Collider and ATLAS Detector

The analysis described in this thesis was performed using data from proton-proton collisions recorded by the ATLAS (A Toroidal LHC ApparatuS) experiment at the Large Hadron Collider (LHC), located at CERN (Conseil Européen pour la Recherche Nucléaire) in Geneva, Switzerland. A brief overview of the LHC and the main components of the ATLAS experiment are given in this chapter.

4.1 The LHC

The LHC at CERN is a 26.7 km two ring superconducting hadron accelerator and collider. It has recently been undertaking Run 2 of its goal to extend the frontiers of particle physics, colliding bunches of up to 10^{11} protons 40 million times per second at unprecedented centre of mass energy of 13 GeV and design instantaneous luminosity of $\sim 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$. The LHC is part of a complex of accelerators, shown in Figure 4.1. Sequentially accelerating the bunches of protons taken from a bottle of hydrogen up to an energy of 450 GeV in the Super Proton Synchrotron (SPS), they are finally accelerated to 6.5 TeV in the LHC ring before being collided at one of the four interaction points. During the 2016 run the LHC achieved and exceeded the design instantaneous luminosity, allowing it to deliver almost $40\,\mathrm{fb}^{-1}$ of data. The potential physics which the LHC will provide in Run 2 is rich, ranging from more precision measurements of standard model (SM) parameters, to further searches for new physics phenomena at the TeV scale, such as supersymmetry (SUSY).

4.2 The ATLAS Experiment

The ATLAS experiment [46] is the largest positioned around the LHC ring. It is a general purpose detector designed to probe p-p collisions, covering the wide range of possible physics at the LHC. The high interaction rates, radiation dose, particle multiplicities and energies impose many requirements upon the design of the detectors for ATLAS, the main components of which are briefly described in the following and shown in Figure 4.2.

ATLAS uses a specific coordinate system and nomenclature to describe the detector and particles emerging from the p-p collisions, where the nominal interaction point (IP) is defined as the coordinate origin and the z-axis defined as the beam direction. The x-y plane is therefore defined transverse to the beam direction, with the positive x-axis pointing from the IP to the centre of the LHC ring and the positive y-axis pointing upwards. The azimuthal angle ϕ is measured around the beam axis and the polar angle θ is the angle from the beam axis. The pseudorapidity is defined as $\eta = -\ln\tan(\theta/2)$ and the distance in η and ϕ space defined as $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$.

The ATLAS experiment consists of a series of concentric cylinders surrounding the beam pipe with a forward-backward symmetry, providing large acceptance in pseudorapidity and almost full azimuthal angle coverage, and can be divided into four major parts: the magnet system, inner detector, calorimeters and the muon system.

4.2.1 Magnet system

The ATLAS detector uses a unique hybrid magnet system comprised of four large superconducting magnets; one solenoid and three toroids, one barrel and two end-caps. The magnet system stores a total energy of 1.6 GJ in a magnetic field volume of $\sim 12000\,\mathrm{m}^3$, with the system 22 m in diameter and 26 m in length.

The central solenoid is positioned within the calorimeter volume with the solenoid windings and LAr calorimeter sharing a vacuum vessel to keep the material thickness as low as possible, ~ 0.66 radiation lengths at normal incidence, such that the calorimeter can achieve the desired performance. It is aligned with the beam axis providing a 2 T axial field for the inner detector which is returned by the steel of the hadronic calorimeter and its girder structure. The solenoid is a single-layer coil wound with high-strength Al-stabilised NbTi conductor housed within a 12 mm thick aluminium support cylinder and has an inner diameter of 2.46 m and outer diameter of 2.56 m, storing 40 MJ of energy.

The barrel toroid is comprised of eight coils encased in individual racetrack-shaped vacuum vessels which surround the calorimeters and end-cap toroids, providing a 0.5 T toroidal field for the muon detectors. The windings are the same in the barrel and end-cap toroids; based on a pure aluminium stabilised Nb/Ti/Cu conductor formed into pancake-shaped coils. The barrel toroid system is supported by the warm structure of aluminium alloy struts mounted between each of the coils. The overall length of the system is 25.3 m, with inner and outer diameters of 9.4 m and 20.1 m respectively and stores a total energy of 1.08 GJ. The end-cap toroids generate the magnetic field required to provide bending power in the

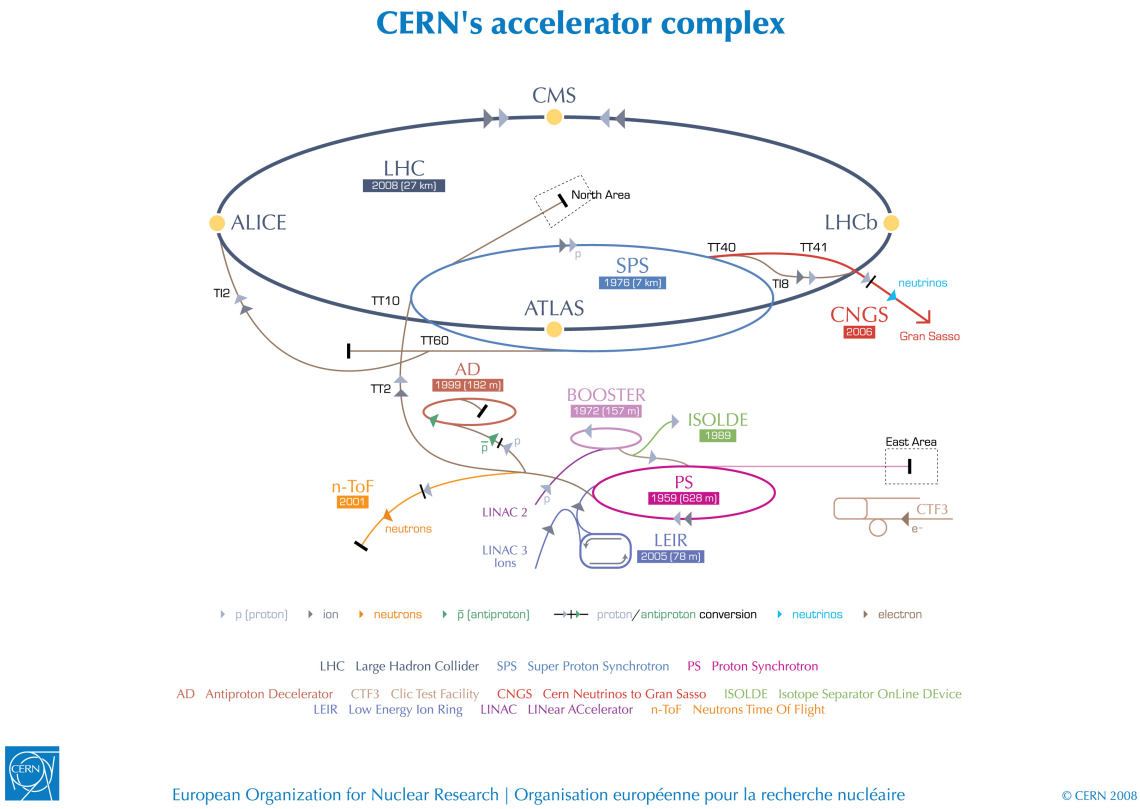


Figure 4.1: The CERN accelerator complex. [45]

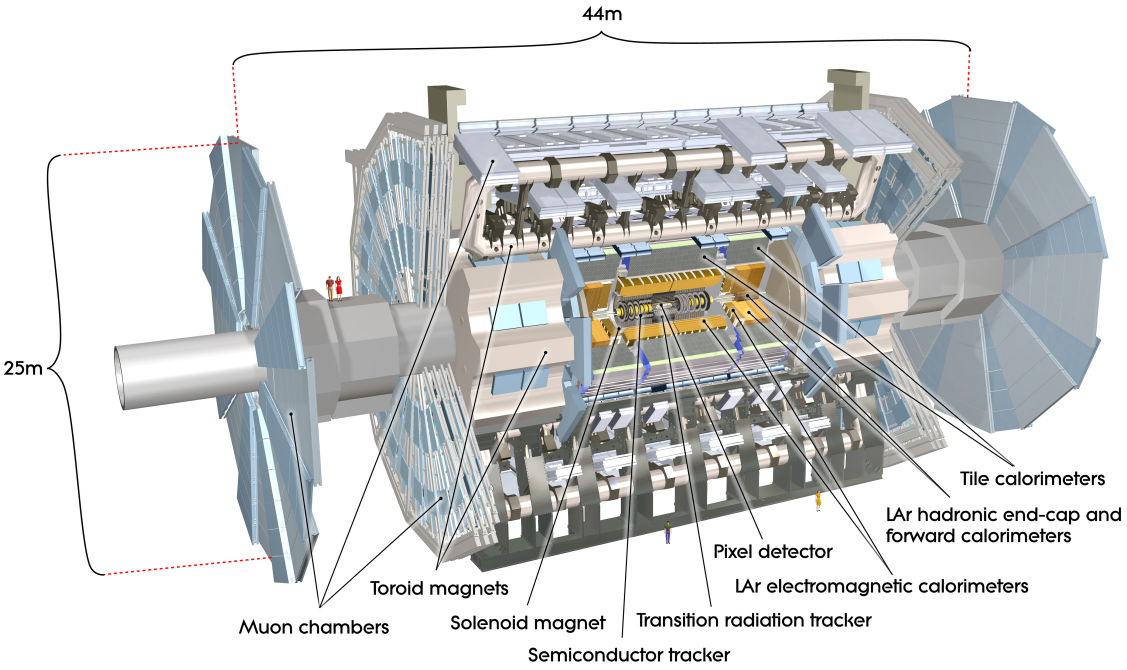


Figure 4.2: Cut away view of the ATLAS detector. [46]

end-cap regions for the muon systems. Each toroid consists of a single cold mass comprised of eight flat, square coil units and eight wedges, which are glued and bolted together to form a rigid structure capable of withstanding the Lorentz forces, they provide 1 T toroidal fields and have a stored energy of 0.25 GJ per end-cap toroid.

4.2.2 Inner Detector

Bunches of protons will collide every 25 ns generating a very high charged track density, with approximately 1000 particles emerging per collision in the pseudorapidity range $|\eta| < 2.5$. The ATLAS inner detector (ID) is designed with four independent but complementary sub-detectors, shown in Figure 4.3, in order to provide the required pattern recognition as well as momentum resolution and primary and secondary vertex measurements. The inner most detector consists of the silicon pixel layers and pairs of silicon microstrips (SCT) which allow for high resolution tracking by providing discrete spatial points in the range $|\eta| < 2.5$. These are arranged in concentric cylinders for the barrel region, while in the end-cap region they are arranged as disks perpendicular to the beam axis. At larger radii there are many layers of gaseous straw tubes interleaved with transition radiation material, known as the transition radiation tracker (TRT). The whole of the ID is immersed in a 2 T magnetic field generated by the central solenoid. During Run 1 of the LHC the performance of the inner detector momentum resolution was measured to be $\frac{\sigma_{p_\mathrm{T}}}{p_\mathrm{T}} = (4.83 \pm 0.16) \times 10^{-4}\,\mathrm{GeV}^{-1} \times p_\mathrm{T}$ [47].

4.2.2.1 Insertable *B*-Layer

The insertable *B*-layer [48] was a new addition for Run 2. It is a fourth high granularity pixel layer closer to the interaction point, at a radius of 3.3 cm from the beam axis, between a new smaller radius beryllium beam pipe and the three original pixel layers, described in Section 4.2.2.2. Its aim is to provide improved tracking, vertexing and *b*-tagging, important for signals involving *b*-jets. The sensors are arranged on 14 staves, each 64 cm long, around the beam pipe with a tilt of 14° which allows for an overlap to provide complete ϕ coverage, however there is no overlap in z to minimise the gap between sensors, $\sim 250\,\mu$m. This arrangement covers the region $|\eta| \leq 2.9$. The inclusion of the IBL is expected to give an improvement for the light jet rejection of the *b*-tagging algorithms by a factor of ~ 1.8–1.9 [49].

4.2.2.2 Pixel detector

As the pixel sensors are positioned in the innermost layers of the detector, they have the highest requirement for leading edge technology to guarantee they meet the stringent specifications on their radiation hardness and resolution. The pixel detector consists of 1744

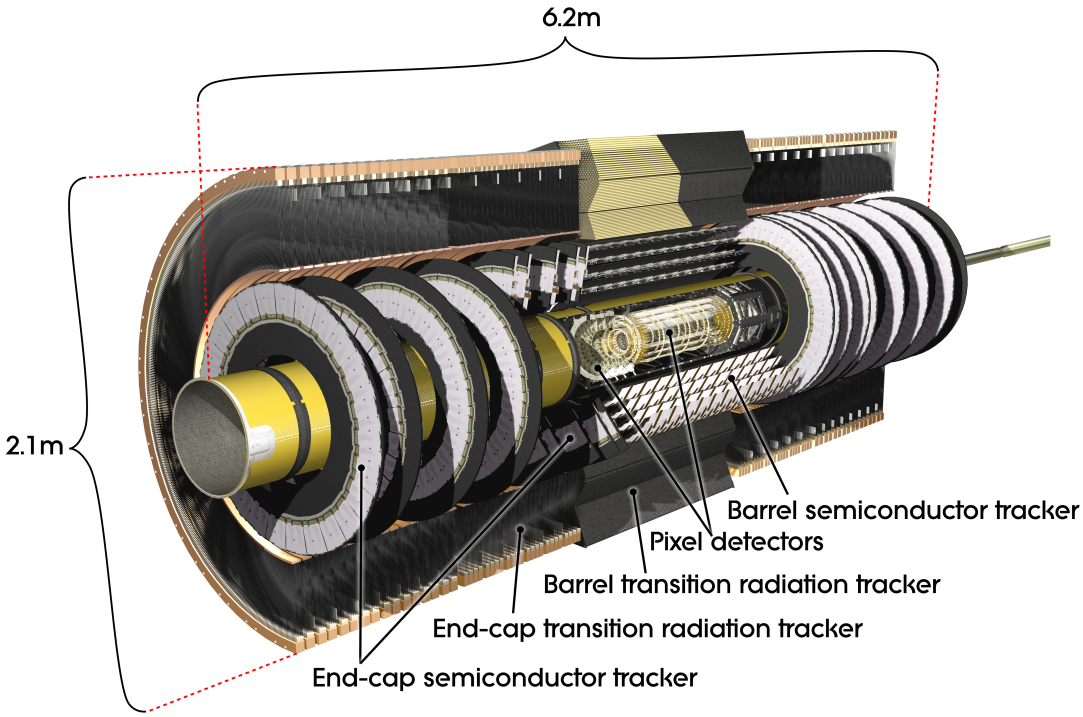


Figure 4.3: Cut away view of the ATLAS inner-detector. [46]

identical pixel sensors made of 250 μm thick oxygenated n-type silicon crystals arranged in three layers in the barrel and in three disks for each of the two end-caps. They are designed to operate with a bias voltage of $\sim 150\,\mathrm{V}$, though an operating voltage of up to 600 V is required to ensure good charge collection as during 10 years of operation, due to radiation damage, the n-type silicon effectively becomes p-type.

4.2.2.3 SCT

The 15912 sensors which make up the SCT use single-sided p-in-n type technology and will also operate initially at a bias voltage of 150 V, again potentially increasing over time to between 250 and 350 V depending on the wear of the detector. Each sensor is $285 \pm 15\,\mu$m thick and has 768 active strips which are 12 cm in length and two additional strips at the bias potential which mark the edges of the sensor.

For the barrel region there are 2112 modules, where each module has 4 sensors, two on the top side and two on the bottom, with an offset of ± 20 mrad about the center of the sensors and mounted on a base board which provides thermal and mechanical structure. These modules are arranged in four coaxial concentric cylindrical layers around the beam axis at a tilt angle of $\sim 11^\circ$ defined by geometrical constraints.

The end-cap has 1976 modules of three different shapes depending on their position within

the SCT end-caps. Each module has two sets of sensors which are glued back-to-back around a base board, again with a relative rotation of ±20 mrad. These modules are then arranged in 9 disk layers for each of the two end-caps.

4.2.2.4 TRT

Designed to detect transition radiation for electron identification the TRT's main components are the 4 mm diameter polyimide drift tubes, made from two films of polyimide and aluminium stuck back-to-back, which are supported by a carbon fibre structure. The tubes are filled with a mixture containing 70% Xe, 27% CO_2 and 3% O_2, which with the cathodes operating at -1530 V gives a gain of 2.5×10^4. For a charged track with transverse momentum, $p_T > 0.5$ GeV and $|\eta| < 2.0$ at least 36 tubes will be traversed.
For the barrel the straws are 144 cm long and the diameter is 31 μm, with tungsten anode wires read out from each end, though only having an active length of ±72 cm means that approximately the middle 2 cm are inefficient. These straws are arranged in three rings of 32 modules each of which consists of a carbon fibre shell with an internal array of polypropylene fibres, acting as the transition radiation material, with the straws embedded within them.
The TRT end-cap has two sets of independent wheels, with 12 wheels of 8 layers closer to the interaction point and 8 wheels of 8 layers further out, with each layer containing 768 straws which are radially oriented with uniform azimuthal spacing. The space between layers is filled with polypropylene radiator foil separated by a polypropylene net.

4.2.3 Calorimeters

The calorimeters cover a large range of $|\eta| < 4.9$ implementing a number of different techniques suited to the vast scope of requirements from the physics processes of interest. The components which make up the calorimeter system are shown in Figure 4.4. The electromagnetic (EM) calorimeter covers the same η range as the inner detector and is required to perform precision measurements of both electrons and photons. The hadronic calorimeter is coarser in granularity but is sufficient to meet the physics requirements for the reconstruction of jets and the measurement and calculation of missing transverse energy, E_T^{miss}.
The depth of the calorimeters was an important design consideration, as the they must provide good containment for EM and hadronic showers, limiting the amount of punch-through into the muon systems. In the barrel the EM calorimeter has a thickness of > 22 radiation lengths, X_0, and in the end caps $> 24X_0$. The total thickness of active calorimeters in the barrel provides approximately 9.7 interaction lengths allowing for good resolution of high energy jets and is sufficient to reduce the amount of punch-through, validated by both simulations and measurements. This thickness, together with the large

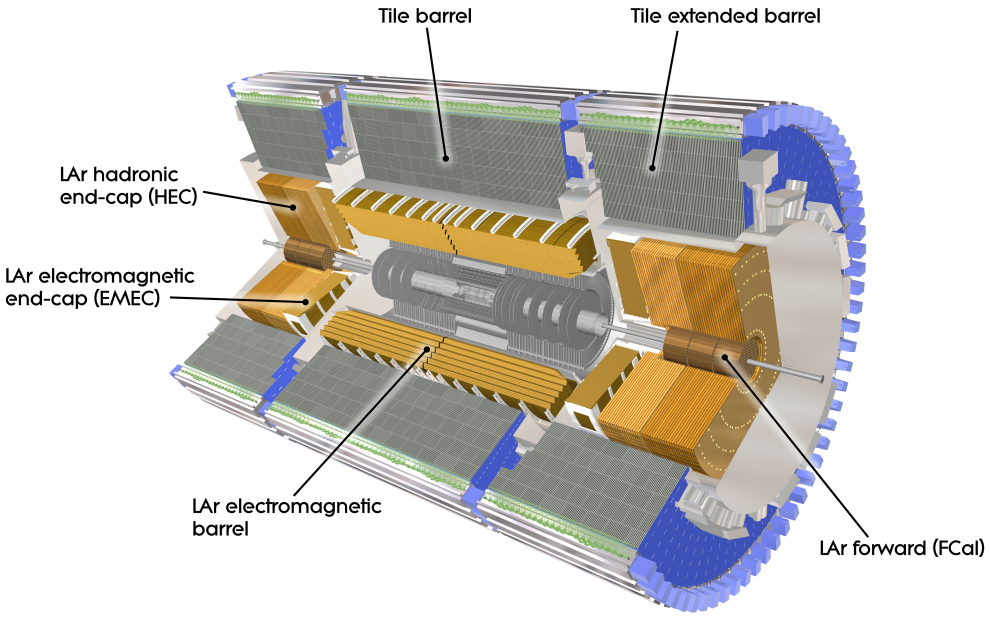


Figure 4.4: Cut away view of the ATLAS calorimeter system. [46]

η coverage, provides a good measurement of E_T^{miss}, which is important for many physics processes including invisible Higgs decays and SUSY searches.

4.2.3.1 Electromagnetic Calorimeter

The electromagnetic calorimeter is divided into three parts, a barrel component covering $0 < |\eta| < 1.475$ and two end-caps covering $1.375 < |\eta| < 3.2$. The calorimeters are constructed of lead-liquid argon detectors where the lead absorbers and electrodes are arranged in an accordion geometry which allows for several active layers in depth, 3 in the region of precision measurements, $0 < |\eta| < 2.5$, and 2 layers in the region $2.5 < |\eta| < 3.2$. The accordion geometry also allows for a detector fully symmetric in ϕ and has no azimuthal cracks. The performance goal of the EM calorimeter for the energy resolution is $\frac{\sigma_E}{E[\text{GeV}]} = \frac{10\%}{\sqrt{E}} \oplus 0.7\%$.

4.2.3.2 Hadronic Calorimeter

The hadronic calorimeter is also divided into three distinct parts; the hadronic tile calorimeter, the liquid argon hadronic end-caps (HEC) and the liquid argon forward calorimeter (FCal). The tile and hadronic calorimeters have been designed to provide an energy resolution of $\frac{\sigma_E}{E[\text{GeV}]} = \frac{50\%}{\sqrt{E}} \oplus 3\%$, while the forward calorimeters provide a resolution of $\frac{\sigma_E}{E[\text{GeV}]} = \frac{100\%}{\sqrt{E}} \oplus 10\%$.

The tile calorimeter is positioned directly outside of the EM calorimeter, with the barrel covering the region $|\eta| < 1.0$ and the end-caps covering $0.8 < |\eta| < 1.7$. It is a sampling calorimeter with scintillating tiles as the active material between steel absorbers, separated into three depth layers of varying depths. Two sides of the scintillating tiles are read-out by wavelength shifting fibres into separate photomultiplier tube (PMTs).
The HEC is behind the EM end-cap calorimeters and so shares the same cryostats. Each end-cap consists of two independent wheels which are positioned such that they overlap with the FCal so that there is no loss of coverage. The wheels are made up of 32 identical wedge shaped modules which are divided into two layers, built of parallel copper plates interleaved with liquid argon (LAr) gaps, where the plates in the layer nearest the center are 25 mm thick whilst those further out are 50 mm.
Recessed about 1.2 m with respect to the EM calorimeter, in order to reduce the amount of neutron albedo where neutrons are scattered back into the inner detector cavity, is the FCal which provides coverage of $3.1 < |\eta| < 4.9$. This obviously limits the depth of the detector and hence necessitates a high density design. The detector is 10 interaction lengths deep, split into three modules in each end-cap, which consist of a matrix of meta plates spacing a series of concentric rods and tubes acting as the electrode structures with LAr in the gaps between. This geometry allows for minimal gaps to avoid pile-up of ions. The first of the three modules is constructed of copper to allow for electromagnetic measurements, while the other two layers are made of tungsten to measure the energy of hadronic interactions.

4.2.4 Muon Spectrometer

The muon spectrometer (MS) forms the outermost part of the detector, shown in Figure 4.5, and is designed to measure the momentum of charged particles leaving the barrel and end-cap calorimeters in the range $|\eta| < 2.7$ and to trigger on these particles in the range $|\eta| < 2.5$. These measurements are based upon the magnetic deflection of these muon tracks in three large superconducting magnets, one for the barrel region providing a field for $|\eta| < 1.4$ and two end-cap magnets bending muon tracks in the region $1.6 < |\eta| < 2.7$, with the bending of muon tracks in the transition region, $1.4 < |\eta| < 1.6$, from a field with contributions from both the barrel and end-cap magnets. The tracks in the barrel region are measured by chambers arranged in three concentric cylindrical shells around the beam axis, while in the transition and end-cap region there are three layers of chambers in a wheel arrangement in planes perpendicular to the beam. The precision measurement of the muon momentum is provided by the muon drift chambers (MDT) which covers the region $|\eta| < 2.7$, except for the forward region which uses cathode-strip chambers (CSC) for the first layer as they have higher rate capability and better time resolution. Complementing the precision tracking chambers are fast trigger chambers, which provide track information about $\sim 10\,\text{ns}$ after a particle goes through, using resistive plate chambers (RPC) in the

barrel and thin gap chambers (TGC) in the end-cap. The performance goal for the muon spectrometer is to provide transverse momentum resolution equivalent to $\frac{\sigma_{p_T}}{p_T} = 10\%$ for muons at $p_T = 1\,\text{TeV}$.

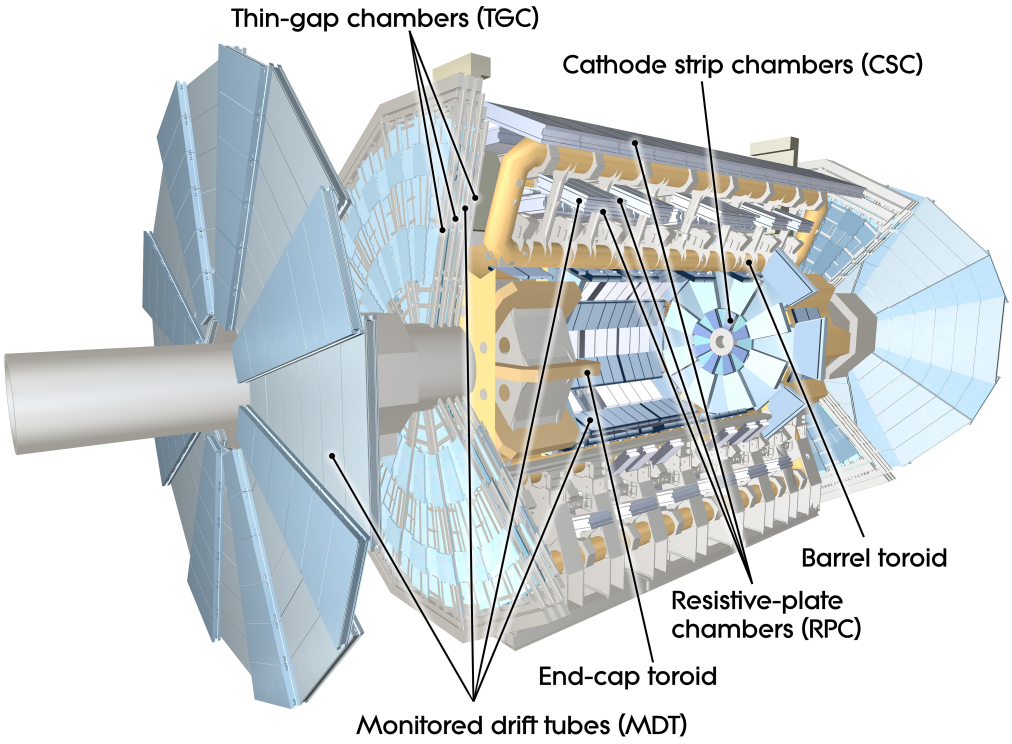


Figure 4.5: Cut away view of the ATLAS muon system. [46]

4.2.4.1 Precision-tracking Chambers

The MDT chambers consist of 3 to 8 layers of pressurised drift tubes which are filled with a mixture of 97% Ar and 3% CO_2 with a central tungsten-rhenium wire at 3080 V collecting electrons from ionisation. This provides precision momentum measurements, achieving a resolution on average of 80 μm per tube or about 35 μm per chamber. The CSC chambers are used in the forward region, $2 < |\eta| < 2.7$, where the particle fluxes and muon track densities are the highest, as these are capable of operating at much higher counting rates, up to $1000\,\text{Hz}\,\text{cm}^{-2}$. They are constructed of multiwire proportional chambers with the wires oriented in the radial direction in four planes, supplying four independent measurements in η and ϕ along each track. This achieves a resolution of 60 μm per CSC plane. In order to achieve the resolutions for these detectors the alignment of the components has to be known to better than 30 μm which is done with a high-precision optical alignment system that relates the position of each chamber to that of its neighbour.

4.2.4.2 Trigger Chambers

Covering the region $|\eta| < 1.05$ the RPCs are gaseous parallel electrode-plate detectors which have good spatial and time resolution in order to provide trigger information. Positioned around the middle and outer layers of the MDT chambers, they consist of two phenolic-melaminic resistive plates separated by 2 mm and filled with a gas mixture of 94.7% $C_2H_2F_4$, 5% Iso-C_4H_{10} and 0.3% SF_6. The electric field between the plates is $\sim 4.9\,\mathrm{kV\,mm^{-2}}$ which allows avalanches to form along the track towards the anode, with the signal being read out by metallic strips mounted on the outer faces of the resistive plates. TGCs cover the forward region, $1.05 < |\eta| < 2.4$, arranged in four layers, three in front and behind the second MDT wheel and the fourth in front of the inner most tracking layer. These chambers operate on the same principle as the multi-wire proportional chambers providing good timing resolution and high rate capability.

4.2.5 Trigger System

The trigger system is an essential part of the ATLAS detector as it is responsible for reducing the huge and unmanageable number of collisions that occur every second to those that contain characteristics which are of interest and may lead to new discoveries. To achieve this a two level trigger system is used; a hardware based Level 1 (L1) trigger which initially reduces the event rate from $\sim 40\,\mathrm{MHz}$ to $\sim 75\,\mathrm{kHz}$ and a High Level Trigger (HLT) reducing the final output rate to $\sim 1\,\mathrm{kHz}$ with an event size of approximately 1.5 MB.

The L1 takes coarse information from the calorimeters and MS, using the L1 calorimeter trigger system (L1Calo) and L1 muon trigger system (L1Muon), and uses fast custom-made electronics to identify high E_T physics objects. From these, regions of interest (RoI) are defined and using the L1 topological modules (L1Topo) topological quantities between L1 objects are constructed, which are then passed to the Central Trigger Processors (CTP) for L1 based selection to be performed. This whole process must be completed within a latency of 2.5 μs before the events which pass are then used to seed the HLT system.

The HLT is a software based trigger system running on a dedicated computer farm that takes the RoI information from L1 and uses fast algorithms to reconstruct and process the events. It has access to finer granularity calorimeter and MS information than the L1 and additionally uses tracking information from the ID. Most HLT triggers utilise a two-stage approach in order to reduce the processing time, where first a fast reconstruction is used to reject the majority of events, before the remaining events undergo a slower precision reconstruction. The average processing time for events is 0.2 s.

Chapter 5

Data and Monte Carlo Samples

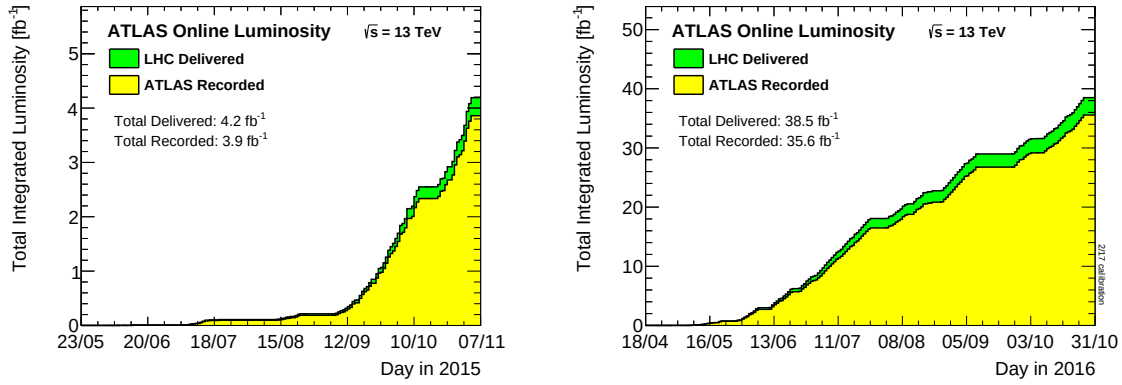
The data samples used in the analysis as well as an overview of the Monte Carlo (MC) generators used to produce simulated samples estimating the signal and background processes is given in the following. All MC samples are passed through the full simulation of the ATLAS detector produced in GEANT4 [50, 51], with both the data and MC samples reconstructed using the same software.

The analysis is dominated by the $t\bar{t}$ process, which constitutes over 90% of the backgrounds. This is split into two similarly sized components; events containing a true hadronically decaying tau and those where a jet fakes a tau. The latter is not well modelled by the MC simulation and so is estimated using a data driven technique, described in Section 8.2, while the former is estimated using MC.
Although relatively small compared to other backgrounds, the production of a Z boson, decaying to a pair of taus, with associated heavy flavour jets[1] has a very similar final state to the signal processes and as such is an important component in the analysis. The MC estimation is used for these processes, however the normalisation is known to not be accurate so a separate control region is used to determine this, described in Section 8.3.
The simulation of backgrounds coming from multi-jet (QCD) processes is particularly challenging, with the MC poorly modelling the data. In the analysis these processes also have final states where a jet fakes the hadronically decaying tau and so are also estimated as part of the data driven estimation described in Section 8.2.
Other non-dominant backgrounds with a small component of the total include single top production, di-boson production (WW, WZ & ZZ) and standard model production of a Higgs boson in association with a Z boson (Zh), which are all estimated from the MC simulations.

[1] Heavy flavour jets refers to those originating from a charm or bottom quark.

5.1 Data samples

The data used for the analysis was collected by the ATLAS detector at the LHC during the 2015 and 2016 running periods, with the LHC operating at a centre-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$ and 25 ns spacing between bunches. The cumulative luminosity versus time delivered by the LHC and recorded by ATLAS for 2015 and 2016 is shown in Figure 5.1. The selected data events were required to have all the relevant components of the ATLAS detector running in good working condition and correspond to an integrated luminosity for 2015 of $3.2\,\text{fb}^{-1}$ and $32.9\,\text{fb}^{-1}$ for 2016, giving a total integrated luminosity of $36.1\,\text{fb}^{-1}$.



(a) (b)

Figure 5.1: The cumulative luminosity versus time delivered by the LHC (green) and recorded by ATLAS (yellow) for 2015 (a) and 2016 (b). The data used for the analysis required all relevant components of ATLAS to be in good working condition and so the luminosity used is slightly less than the total recorded.

5.2 Monte Carlo Samples

Simulated samples are used to model the resonant di-Higgs production from the extended Higgs sector 2HDM heavy scalar Higgs, H, and the bulk RS graviton, G^*_{KK}, in extra Kaluza-Klein dimensions. 14 mass points (260, 275, 300, 325, 350, 400, 450, 500, 550, 600, 700, 800, 900, 1000 GeV) were generated for the H and a subset of 9 mass points (260, 300, 400, 500, 600, 700, 800, 900, 1000 GeV) were generated for the G^*_{KK}, with $m_h = 125\,\text{GeV}$ for the SM Higgs bosons produced in the decays. The H events were produced with a narrow width of $\Gamma = 4\,\text{MeV}$. The width of the G^*_{KK} signal events is dependent on the coupling constant $c = k/M_{\text{Pl}}$ where k is the parameter corresponding to the curvature of the warped extra dimension and $M_{\text{Pl}} = 2.4 \times 10^{18}\,\text{GeV}$ is the effective four-dimensional Planck scale. The benchmark model used has $c = 1.0$ which corresponds to a width $\Gamma = 55\,\text{GeV}$ assuming a graviton mass of 1 TeV. These signals are simulated using an effective field theory (EFT) model implemented in MadGraph5_aMC@NLO v2.2.2 [52] at leading order (LO) and interfaced to Pythia 8.186 [53] for parton showering. For both signal models

the A14 tune [54] is used together with the NNPDF2.3LO parton distribution (PDF) set [55].

For the non-resonant di-Higgs production, a Higgs boson mass of $m_h = 125.09\,\text{GeV}$ and SM production diagrams are assumed, with events being simulated using an EFT model implement in MadGraph5_aMC@NLO v2.2.2 [56] at next-to-leading order (NLO). For parton showering and hadronisation this is interfaced to Herwig++ [57] and the UEEE5 tune together with the CTEQ6L1 PDF set [58] are used.

To simulate events for the $t\bar{t}$ and, single-top Wt and s-channel production processes the Powheg-Box v2 generator [59, 60, 61] is used, with the CT10 PDF set [62] used in the matrix element calculations. For the single-top electroweak t-channel process events are generated using Powheg-Box v1, which uses the fixed four-flavour PDF set CT10f4 for NLO matrix element calculations in the four-flavour scheme. The top-quark spin correlations are preserved for all top processes, with the top-quarks decayed using MadSpin [63] for the t-channel production. Pythia 6.428 [64] together with the CTEQ6L1 PDF sets with the corresponding Perugia 2012 (P2012) tune [65] are used to simulate the parton shower, fragmentation and the underlying event. EvtGen v1.2.0 [66] is used to model the properties of bottom and charm hadron decays. For all events $m_t = 172.5\,\text{GeV}$. For the single-top processes the generator NLO cross-sections are used, while for the $t\bar{t}$ production the cross-section is calculated at NNLO+NNLL [67].

The production of $W^{\pm}$ or Z^0 bosons with associated jets is simulated using the Sherpa 2.2.1 generator [68]. The Comix [69] and OpenLoops [70] matrix element generators are used to calculate matrix elements for up to two partons at NLO and four partons at LO and are then merged with the Sherpa parton shower [71] following the ME+PS@NLO prescription [72]. The predicted cross-sections using NNLO calculations are used to normalise the events from $W^{\pm}/Z^0$+jets processes, with the CT10NLO PDF set used in conjunction with a dedicated parton shower tuning developed by the Sherpa authors.

The Sherpa 2.2.1 generator [68] is used to simulate processes where di-bosons are produced, with one decaying leptonically and the other decaying hadronically. They are calculated for up to three additional partons at LO and up to one or zero additional partons at NLO for ZZ and WW/WZ respectively. The CT10 PDF set used in conjunction with the dedicated parton shower tuning developed by the Sherpa authors and the NLO cross-sections are taken from the generator.

Associated Higgs production with a Z^0 boson that subsequently decays to a $bb\tau\tau$ final state is an irreducible background to this analysis. Pythia 8.186 is used to generate the

processes $qq \to Z^0h(Z^0 \to \tau\tau, h \to bb)$ and $qq \to Z^0h(Z^0 \to bb, h \to \tau\tau)$, together with the NNPDF2.3LO PDF set and the A14 tune. Due to difficulties in modelling the gluon-fusion initiated counterpart to the $qq \to Z^0h(Z^0 \to bb, h \to \tau\tau)$, its cross-section is scaled up by 6% to account for the absent gluon-fusion component. Powheg-Box v2 [73] along with the CT10 PDF sets are used to the generate the $gg \to Z^0h(Z^0 \to \tau\tau, h \to bb)$ process. For the modelling of non-perturbative effects the AZNLO tune [74] is used with the CTEQ6L1 PDF set. The simulation of the parton shower, fragmentation and the underlying event uses Pythia 8.186. For all Z^0h processes the modelling of the bottom and charm hadron decays is done with EvtGen v1.2.0. The contributions from all other SM Higgs processes were checked and found to be negligible.

Chapter 6

Object Reconstruction

An overview of the definitions of reconstructed physics objects is given in the following. These reflect the recommendations of the ATLAS combined performance groups who provide algorithms and tools for reconstruction and identification.

6.1 Jets

In particle physics the concept of a jet refers to the 'cone-shaped' collimated collection of hadrons produced by the hadronisation and fragmentation of a quark or gluon originating from the p–p collision. Jet candidates are reconstructed from topological clusters of energy deposited in the calorimeter cells using the anti-k_t algorithm [75] with a radius parameter $R = 0.4$. The jet energies are determined from clusters initially reconstructed at the EM scale which are then corrected with scale factors derived with MC and data such that the energy deposits of hadrons and their corresponding resolutions are correctly reconstructed, detailed in Ref. [76]. The jet vertex tagging (JVT) algorithm [77] is used to identify and select jets from the hard-scatter interaction using tracking and vertexing information and removing energy deposits originating from pileup interactions, detector effects and cosmic rays. Simulated jets from MC are labelled depending on which hadrons are found within $\Delta R = 0.3$ of the centre of the jet with $p_{\mathrm{T}} > 5\,\mathrm{GeV}$, with those containing b-hadrons labelled as b-jets, those with no b-hadrons but c-hadrons labelled as c-jets and are otherwise labelled as light-jets. All jets are required to have $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ and $|\eta| < 2.4$.

6.1.1 b-tagging

In order to identify jets originating from b-quarks the MV2 multivariate discriminant [78][79] is used. This exploits the properties of B-hadrons compared to light quark hadrons, such as the secondary (and even tertiary) vertices produced within jets due to the b-hadron lifetime being sufficiently long that they travel a few mm before decaying. The relatively large b-hadron mass also allows decays to higher multiplicities of particles within the jet with

higher transverse momentum leading to larger invariant jet mass. The algorithm combines variables such as these with variables from basic b-tagging algorithms, such as track-based impact parameter taggers IP2D and IP3D or vertex based taggers SV1 and JetFitter, into a BDT trained against a background sample containing 10% c-quark initiated jets. The working point used has an average tagging efficiency of 70% for b-jets in $t\bar{t}$ events and provides rejection factors of 381 and 12 for light- and c-jets respectively.

The simulated events are corrected to compensate for mismodelling between data and MC in the b-tagging efficiency by applying scale factors for the b-jets derived in di-leptonic $t\bar{t}$ events. These are consistent with unity within a few percent over the majority of the jet p_{T} range.

As b-jets contain B-hadrons and these undergo semi-leptonic decays, the b-jets are corrected for the momentum carried away by muons in such cases. This is done by adding the four-momentum of the closest muon in ΔR found within a jet to the calorimeter based jet energy after removing calorimeter energy deposits associated with the muon. A p_{T} dependent scale factor is also applied to the jet four-vector to account for biases in the response and correct for missing momentum related to neutrinos produced in the semi-leptonic b-hadron decays, in order to correct the central value and improve the resolution of the di-b-jet mass distribution. The techniques used are described in Ref. [80].

6.1.2 Truth-tagging

Non-dominant backgrounds estimated from MC with low b-jet multiplicity or that are not typically produced in association with b-jets have very few events which survive the 2 b-tag requirement of the signal region, resulting in badly modelled backgrounds due to the poor statistical power of the remaining samples. To improve this a 'truth-tagging' technique is used. For a given jet this involves sampling a random efficiency from a cumulative distribution built from tagging efficiencies above the 70% working point and assigning that jet the corresponding MV2c10 score to this efficiency. The efficiencies used to build the distribution are parameterised as a function of the jet flavour, transverse momentum and pseudorapidity, as well as the process type. After the generation of the MV2c10 values for the jets, the event is weighted by the efficiency of the selected jets to actually pass the b-tagging requirements used in the analysis. In contrast to direct tagging, where jets failing the selection based on the b-tagging discriminant are removed, all events are kept retaining the full statistical power of the sample. This process is applied to all samples except the signals, $t\bar{t}$, single top and $Z \to \tau\tau + bb$ processes.

6.2 Electrons

Electrons leave tracks in the ID and produce showers of particles in the EM calorimeter where they deposit their energy through Bremsstrahlung radiation. Electron candidates are required to have $p_{\mathrm{T}} > 7\,\mathrm{GeV}$, and $|\eta| < 2.47$, calculated based on the direction of the track and cluster of the electron and the energy deposited in the calorimeter. Candidates in the transition region between the calorimeter barrel and endcap, $1.37 < |\eta| < 1.52$, are vetoed. They are identified and selected through requirements on the shape of clusters of energy deposits in the calorimeter, measured track properties and the associated track quality from the ID, and track-to-cluster matching, with identification using a likelihood technique [81] required to pass the 'loose' working point. This, in combination with additional track hit requirements, provides an electron identification efficiency of 95%. Isolation requirements are also imposed which require no nearby tracks or calorimeter energy deposits within a p_{T}-dependent variable cone around the electron candidate. Specifically the 'loose' isolation working point is required, achieving 99% electron efficiency, which is independent of the electron p_{T}. Control regions for the estimation of background are provided by inverting this isolation requirement.

6.3 Muons

Muons are characterised through there ability to pass through the ID and calorimeters to the outer MS detectors, leaving tracks in the ID and MS and a small amount of energy in the calorimeters. This is due to their relatively large mass compared with electrons meaning they undergo less Bremsstrahlung radiation in the detector, depositing less energy. Muon candidates are required to have $p_{\mathrm{T}} > 7\,\mathrm{GeV}$ and $|\eta| < 2.7$, and are defined from tracks reconstructed separately in the ID and MS. These are required to have a minimum number of hits in each system and be geometrically and momentum compatible. Information from both the ID and MS are used in a combined fit to improve the momentum measurement of the muon and a likelihood based method [82] is used to define identification working points, with candidates required to pass the 'loose' criteria. An equivalent 'loose' isolation working point is required to reject non-prompt muons in the tracking detectors and calorimeters, achieving a 99% efficiency which is p_{T} independent. As with electrons, the inverse of the isolation requirements are used to define control regions for background estimation.

6.4 Hadronic Taus

Candidates for hadronically decaying taus are reconstructed using jets formed using the anti-k_t algorithm, with a radius parameter $R = 0.4$, and clusters of calorimeter cells, calibrated using local hadronic calibrations, to seed the visible hadronic tau ($\tau_{\text{had-vis}}$) decay

reconstruction algorithm [83][84][85]. These are required to have $p_{\mathrm{T}} > 10\,\mathrm{GeV}$ and $|\eta| < 2.4$, with candidates in the calorimeter barrel-endcap transition region, $1.37 < |\eta| < 1.52$, being vetoed. They are categorised by the number of charged pions in the $\tau_{\text{had-vis}}$ decays, referred to as 1- and 3-track (or prong) decays. The reconstruction algorithm provides poor discrimination for identification purposes against the jet background by itself so a multi-variate BDT discriminant is used to provide good separation between jets and $\tau_{\text{had-vis}}$. The BDT makes use of variables based on information about the energy deposits in the calorimeter around the $\tau_{\text{had-vis}}$ candidate as well as the tracks, with separate BDTs trained for the 1- and 3-prong candidates. The BDT score is transformed such that the true taus are evenly distributed as a function of BDT score independent of their p_{T} and the pile-up conditions in the event. It is also necessary to distinguish between electrons and 1-prong taus so candidates with one core track that overlap within $\Delta R < 0.4$ of an electron candidate, required to have $p_{\mathrm{T}} > 5\,\mathrm{GeV}$ and high electron identification likelihood score, are vetoed. The selected $\tau_{\text{had-vis}}$ candidates are required to have one or three prong decays, $p_{\mathrm{T}} > 20\,\mathrm{GeV}$, $|\eta| < 2.5$, unit charge and pass the 'medium' ID BDT working point.

6.4.1 Fake Taus

To provide regions enriched in events where a jet fakes the $\tau_{\text{had-vis}}$, an anti-ID-τ selection is defined. These are required to have failed the 'medium' ID requirement but have a BDT score above a 0.35 threshold such that the properties of the jets are reasonably similar to those of a hadronic tau and ensures the composition of quark and gluon initiated jets resembles that of the signal region.

6.5 Missing Transverse Energy

The missing transverse momentum, $p_{\mathrm{T}}^{\mathrm{miss}}$, uses an object based definition [86] and is the negative of the vector sum of the p_{T} of the fully calibrated and reconstructed physics objects described above. An additional Track Soft Term (TST) is included to account for transverse momentum deposited in the detector not associated to any physics objects and is calculated from ID tracks matched to the PV with $p_{\mathrm{T}} > 400\,\mathrm{MeV}$ to make it relatively insensitive to contributions from pile-up interactions. The missing transverse energy, $E_{\mathrm{T}}^{\mathrm{miss}}$, is the magnitude of the missing transverse momentum vector. For events containing an anti-ID-τ used for background estimation, the anti-ID-τ's are treated as if they were τ objects in the calculations of the $E_{\mathrm{T}}^{\mathrm{miss}}$ and $p_{\mathrm{T}}^{\mathrm{miss}}$ to ensure they are calibrated at the tau rather than jet energy scale.

6.6 Overlap Removal

Signals in the detectors can often get reconstructed by the separate algorithms as more than one physics object, as described above. An overlap removal (OR) procedure is applied between objects which have a geometric overlap in ΔR below a certain threshold to avoid double counting. The decision for which object takes priority is based on the ΔR thresholds and in some cases additional requirements are applied to determine which object is kept. These are detailed in Table 6.1 with the precedence of the objects represented by their order in the table.

Overlapping Objects	$\Delta R <$	Additional Requirements	Priority
e-jet	0.2	—	e
jet-e	0.4	—	jet
μ-jet	0.4	$n_{tracks}(p_T^{track} > 500\,\text{MeV}) < 3$	μ
jet-μ	0.4	$n_{tracks}(p_T^{track} > 500\,\text{MeV}) \geq 3$	jet
e-μ	0.2	—	μ
μ-τ	0.2	—	μ
e-τ	0.2	—	e
τ-jet	0.2	τ ID working point and b-tagging score	1) BDT 'medium' τ 2) b-tagged jet 3) anti-ID-τ 4) c-/light-jet

Table 6.1: Procedure for removing overlapping objects based on the ΔR threshold values and any additional requirements. The object to which priority is given to is kept.

6.7 Missing Mass Calculator

The invariant mass of the di-τ system is reconstructed using the Missing Mass Calculator (MMC) [87] algorithm which attempts to provide a full reconstruction of the event topology. This requires solving equations for 7 unknowns for events containing one leptonically decaying tau, τ_{lep}, and one hadronically decaying tau, τ_{had}, in the kinematically allowed phase space to correct for the momentum carried away by the neutrinos from the tau decays. The unknowns are the x-, y- and z- components of the momentum of the neutrino(s) from each tau and the invariant mass of the di-neutrino system ($\nu\nu$) for the τ_{lep}. The kinematic

constraints related to τ_i in terms of the visible decay products and the neutrinos are

$$\begin{aligned} m^2_{\tau,i} &= \left(p_{\ell,i} + p_{\nu\nu,i}\right)^2, \\ \vec{p}^{\,2}_{\nu\nu,i} &= E^2_{\nu\nu,i} - m^2_{\nu\nu,i}, \end{aligned} \tag{6.1}$$

for the τ_{lep}, and

$$m^2_{\tau,i} = \left(p_{\tau,i} + p_{\nu,i}\right)^2, \tag{6.2}$$

for the τ_{had}. This information is not sufficient to find an exact solution as the system is under-constrained, so additional knowledge about the tau decay kinematics is included to determine the more likely solutions. The algorithm scans variables related to the 4-vectors of the $E_{\text{T}}^{\text{miss}}$ and the visible decay products of two taus, using the kinematic constraints to reconstruct the momentum of each neutrino and weighting each scan point by the probability obtained. The estimator for the mass of the di-τ system is then defined from the maximum of the likelihood function constructed from the scan.
The performance of the MMC is highly correlated to the $E_{\text{T}}^{\text{miss}}$ resolution, which is also included in the scan such that the MMC can be adjusted for possible mismeasurements of the $E_{\text{T}}^{\text{miss}}$. The MMC achieves a resolution for the di-tau system of 16% and has a high efficiency ($> 99\%$) to find a solution for the signal and $Z \to \tau\tau$ processes, an improvement over other methods. Backgrounds which do not contain a di-τ resonance, for example $t\bar{t}$, have a lower efficiency providing additional rejection against them.

Chapter 7

Event Selection

The trigger and preselection requirements for the analysis are described in this chapter followed by the details of the BDTs used for the analysis, which the other MVA methods discussed later will be benchmarked against. The analysis utilises two sets of triggers: single lepton triggers (SLT) and lepton-plus-tau triggers (LTT). These are treated and optimised as separate channels as the background composition and modelling differ in the two channels due to different acceptance cuts. The SLT provides the majority of the events and is the most sensitive for the SM non-resonant analysis and in general is most sensitive across the entire mass range for the resonant analysis. The LTT provides similar sensitivity to the SLT in the low mass region of the resonant analysis, and although less sensitive elsewhere provides some gain to using solely the SLT.

7.1 Trigger

Events are tested to see if they pass a number of triggers chosen based on the final state topology. For each run period the triggers are required to have the lowest p_{T} thresholds, to get maximum efficiency, and be unprescaled[1], to improve the MC modelling of the data. Firstly as one of the taus decays leptonically for the signal this can be exploited by requiring that events pass one of the available SLT split into separate electron and muon channels, summarised in Table 7.1. The nomenclature related to the SLT are of the format *HLT*, followed by the light lepton type (e/mu) and its corresponding p_{T} threshold, identification requirement (prefixed with *lh*) and finally the isolation requirement (prefixed with *i*). Electron triggers with the *nod0* suffix do not use the $d0$ transverse impact parameter information in the electron likelihood-based identification. The lowest threshold 2015 electron trigger name also contains the L1 EM trigger from which it was seeded, *L1EM20VH*, where the minimum threshold is $p_{\mathrm{T}} > 20\,\mathrm{GeV}$, *V* denotes the p_{T} threshold

[1]Prescaling refers to the process of randomly vetoing events that pass a trigger to reduce overall trigger rates.

changes as a function of η to account for energy loss and H denotes hadronic core isolation requirements are applied.

- For electrons several triggers are used in an OR configuration to maximise acceptance:
  - Events are required to have at least one electron with $p_{\mathrm{T}} > 24\,\mathrm{GeV}$ passing the 'medium' identification criteria and the 'loose' isolation requirements (the p_{T} threshold is raised to 26 GeV and the identification required to be 'tight' in later data-taking periods), OR
  - at least one electron with $p_{\mathrm{T}} > 60\,\mathrm{GeV}$ and 'medium' identification criteria but no isolation requirement, OR
  - at least one electron with $p_{\mathrm{T}} > 120(140)\,\mathrm{GeV}$, depending on the run period, satisfying the 'loose' identification criteria.
- For the muon channel, events must have a least one muon with $p_{\mathrm{T}} > 24(26)\,\mathrm{GeV}$, depending on the data-taking period, which are required to satisfy the 'medium' isolation requirements.

Run Periods	Electron Triggers	Muon Triggers
2015	HLT_e24_lhmedium_iloose_L1EM20VH ||	
	HLT_e60_lhmedium ||	HLT_mu24_imedium
266904 – 284484	HLT_e120_lhloose	
2016	HLT_e24_lhmedium_ivarloose ||	
A - D3	HLT_e60_lhmedium_nod0 ||	HLT_mu26_imedium
296939 – 302872	HLT_e140_lhloose_nod0	
2016	HLT_e26_lhtight_ivarloose ||	
D4 onwards	HLT_e60_lhmedium_nod0 ||	HLT_mu26_ivarmedium
302919 – 310216	HLT_e140_lhloose_nod0	

Table 7.1: Single lepton triggers (SLT) used for each data-taking period of the analysis with corresponding run numbers. The electron triggers are used in an OR configuration.

Events that fail the SLT are then considered against the LTT. The LTT requires at least one light lepton and at least one hadronic tau. The hadronic tau must have $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ and satisfy the 'medium' identification criteria. The electron and muon are required to pass the $p_{\mathrm{T}} > 17\,\mathrm{GeV}$ and $p_{\mathrm{T}} > 14\,\mathrm{GeV}$ thresholds respectively, with both required to meet the 'medium' identification requirements. For later data-taking periods the 'loose' isolation criteria must additionally be satisfied. During 2016 data-taking the LTT had an additional Level-1 requirement for a jet passing a $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ threshold. These triggers are summarised in Table 7.2 and follow the nomenclature *HLT*, followed by the light lepton

type (e/mu) and its corresponding p_{T} threshold, identification requirement (prefixed with lh) and isolation requirement (prefixed with i), then tau followed by the hadronic tau p_{T} threshold, identification requirement and minimum number of tracks in the tau region of interest.

Run Periods		Triggers
2015 & 2016 A	$e+\tau$	HLT_e17_lhmedium_nod0_tau25_medium1_tracktwo
$266904-300287$	$\mu+\tau$	HLT_mu14_tau25_medium1_tracktwo
2016 B onwards	$e+\tau$	HLT_e17_lhmedium_nod0_ivarloose_tau25_medium1_tracktwo
$300345-310216$	$\mu+\tau$	HLT_mu14_ivarloose_tau25_medium1_tracktwo

Table 7.2: Events failing the single lepton triggers are tested against lepton-plus-tau triggers (LTT). The triggers used for each data-taking period of the analysis are listed with corresponding run numbers. For the 2016 data-taking period the LTT had an additional Level-1 required for a jet with $p_{\mathrm{T}} > 25\,\mathrm{GeV}$.

If an event contains jets that have not been associated with any properly reconstructed energy deposits the event is discarded, the variables and criteria used to veto events can be found in [88]. These can be caused by beam induced backgrounds, cosmic-ray showers or hardware problems such as calorimeter noise. Events are also required to contain at least one primary vertex, defined as a point of proton-proton interaction in the luminous region associated with at least two tracks of $p_{\mathrm{T}} > 400\,\mathrm{MeV}$. If more than one primary vertex is present in an event, the one with highest summed p_{T}^2 is considered.

7.2 Preselection

Events passing the above trigger selection and that are compatible with a final state containing at least one light lepton, at least one hadronic tau, at least two jets and $E_{\mathrm{T}}^{\mathrm{miss}}$ are then subject to the preselection criteria described in the following. For the SLT events must contain exactly one electron which passes the criteria for the 'tight' identification working point or one muon passing the 'medium' identification criteria with the additional requirement that $|\eta| < 2.5$. These light leptons are required to have transverse momentum that is 1 GeV higher than the trigger threshold corresponding to the data-taking period, 25 GeV or 27 GeV, such that the trigger efficiency is (close to) on plateau where the MC models the data well. There must be exactly one 'medium' identification hadronic tau with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ and $|\eta| < 2.3$. The event is also required to contain at least two jets with $p_{\mathrm{T}} > 45\,\mathrm{GeV}$ and $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ for the leading and sub-leading jets respectively.
The LTT events must contain exactly one electron which meets the 'tight' identification criteria with $18\,\mathrm{GeV} < p_{\mathrm{T}} < 25(27)\,\mathrm{GeV}$, or one 'medium' identification muon with $15\,\mathrm{GeV} < p_{\mathrm{T}} < 25(27)\,\mathrm{GeV}$, where the upper cut on transverse momentum corresponds to

the SLT thresholds applied for the equivalent data-taking period. There must be exactly one hadronic tau with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$, such that the trigger efficiency is on plateau, and $|\eta| < 2.3$. There must be at least two jets with $p_{\mathrm{T}} > 80\,\mathrm{GeV}$ and $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ for the leading and sub-leading jet respectively, where the higher leading jet transverse momentum is due to a Level-1 Trigger requirement of a jet in the event.
In addition to the above requirements, the light lepton and hadronic tau in the event must have opposite sign, $Q = |1|$ charges. The invariant mass of the di-tau system is reconstructed using the MMC algorithm, described in Sec. 6.7, and is required to be $m^{\mathrm{MMC}}_{\tau_\ell\tau_h} > 60\,\mathrm{GeV}$ as it was found that below this value the data-MC agreement was poor due to the algorithm failing.
The events are categorised by the number of b-tagged jets in the event, where the events with exactly two b-tags are used in the multivariate analysis, defining the signal region. The expected yields for signal and background events in this region as well as number of data are given in Table 7.3. The zero and one b-tag categories are used as control and validation regions for the background estimation, where in the zero b-tag region the highest p_{T} jets are considered and in the one b-tag region the b-tagged jet and highest p_{T} non-tagged jet are used. Events with more than 2 b-jets are vetoed to ensure orthogonality to the $hh \to bbbb$ analysis.

Sample	SLT	LTT
$G(m_G = 300\,\text{GeV},\, c = 1.0)$	23.4 ± 0.4	7.21 ± 0.20
$G(m_G = 500\,\text{GeV},\, c = 1.0)$	43.6 ± 0.4	10.1 ± 0.2
$G(m_G = 800\,\text{GeV},\, c = 1.0)$	7.98 ± 0.07	1.11 ± 0.02
$H(m_H = 300\,\text{GeV})$	9393 ± 105 (39.7 ± 0.4)	2768 ± 59 (11.7 ± 0.2)
$H(m_H = 500\,\text{GeV})$	10830 ± 115 (3.50 ± 0.04)	2760 ± 58 (0.89 ± 0.02)
$H(m_H = 800\,\text{GeV})$	1207 ± 11 (0.149 ± 0.001)	185 ± 4 (0.0228 ± 0.0005)
Non-res. SM $hh \to \tau_\ell \tau_h bb$	1.07 ± 0.04	0.28 ± 0.02
$t\bar{t}$	19353 ± 95	1492 ± 26
Single top (Wt)	964 ± 19	62.7 ± 4.8
Single top t-channel	149 ± 8	6.18 ± 1.45
Single top s-channel	9.65 ± 1.04	0.46 ± 0.14
DY	2.86 ± 0.94	0.25 ± 0.12
$Z^0/\gamma \to \tau\tau + bb$	430 ± 16	116 ± 4
$Z^0/\gamma \to \tau\tau + bc$	23.7 ± 1.2	6.91 ± 0.46
$Z^0/\gamma \to \tau\tau + bl$	8.82 ± 0.51	2.79 ± 0.13
$Z^0/\gamma \to \tau\tau + cc$	11.2 ± 0.5	3.85 ± 0.19
$Z^0/\gamma \to \tau\tau + cl$	2.60 ± 0.07	0.85 ± 0.03
$Z^0/\gamma \to \tau\tau + l$	0.66 ± 0.01	0.213 ± 0.004
$Z^0/\gamma \to \ell\ell + bb$	114 ± 5	$4.61 \pm .065$
$Z^0/\gamma \to \ell\ell + bc$	6.32 ± 0.50	0.28 ± 0.04
$Z^0/\gamma \to \ell\ell + bl$	2.46 ± 0.11	0.089 ± 0.045
$Z^0/\gamma \to \ell\ell + cc$	2.27 ± 0.43	0.097 ± 0.036
$Z^0/\gamma \to \ell\ell + cl$	0.94 ± 0.19	0.030 ± 0.003
$Z^0/\gamma \to \ell\ell + l$	0.29 ± 0.01	0.008 ± 0.001
W+jets	8.16 ± 1.23	0.61 ± 0.15
Fake τ backgrounds	7691 ± 47	446 ± 19
WW	1.41 ± 0.21	0.070 ± 0.026
WZ	6.96 ± 0.53	0.52 ± 0.11
ZZ	11.0 ± 0.6	2.15 ± 0.29
$Zh(Z \to \ell\ell, h \to bb)$	4.78 ± 0.09	0.87 ± 0.04
Total SM Background	28796 ± 109	2147 ± 33
Data	28613	2171

Table 7.3: The expected number of signal and backgrounds events from Monte Carlo with their statistical uncertainties and observed number of data after applying the preselection criteria and requiring exactly two b-tagged jets. The yields for the 2HDM heavy scalars setting $\tan\beta = 2$ are shown in brackets as the yields for the benchmark model are found to be unrealistic and correspond to a previously excluded region of phase-space.

7.3 Multi-Variate Analysis Training

An MVA method is used to improve separation between the signal and background. The baseline for the ATLAS publication analysis is to use a series of BDTs, described in the following, which provide a benchmark against which additional methods such as SVMs can be compared.

7.3.1 Variables

The inputs to the MVAs are variables considered to provide reasonable discrimination between the signal and background. Some variables may exhibit separation power individually, however this is not always the case and correlations between variables can be exploited to separate between signal and background. By sequentially adding or removing individual variables, retraining the method and observing the effect on the FOM, the importance of each variable can be assessed.

In the case of the BDT TMVA ranks the variables by looking at how frequently they are used to split at nodes and weights them by the square of the separation gained (the Gini Index for the benchmark trainings) and by the number of events in the node. The list of inputs was then reduced by removing the lowest ranked variables and retraining, checking the impact on the Receiver Operating Characteristic (ROC) curves, which show the signal efficiency versus background rejection for given cuts on the classifier outputs, and the expected limits themselves. Variables were removed until a decrease in performance was observed. The variables used in the benchmark trainings are:

- m_{bb}: The invariant mass of the di-b-jet system.
- $m_{\tau_\ell\tau_h}^{\mathrm{MMC}}$: The invariant mass of the di-tau system calculated using the MMC.
- m_{hh}: The invariant mass of the di-Higgs system reconstructed from the di-b-jet and di-tau systems. The four-momenta of the di-b-jet and di-tau systems are scaled by m_h/m_{bb} and $m_h/m_{\tau_\ell\tau_h}^{\mathrm{MMC}}$ respectively to improve the resolution, where $m_h = 125\,\mathrm{GeV}$ is the value used in the simulation.
- $\Delta\phi\,(h,h)$: The $\Delta\phi$ angular separation between the two reconstructed 125 GeV Higgs candidates, where the directions are taken from the di-b-jet system directly and from the MMC fit for the di-tau system.
- $\Delta R\,(l,\tau)$: The ΔR between the visible decay products of the taus.
- $\Delta R\,(b_1,b_2)$: The ΔR between the two b-jets.

- $\Delta p_{\mathrm{T}}\,(\ell,\tau)$: The transverse momentum imbalance between the lepton and the visible decay products of the hadronic tau, exploiting the difference in the number of neutrinos in leptonic and hadronic tau decays.

- m_{T}^{W}: The transverse mass of the lepton and missing transverse momentum calculated as

$$m_{\mathrm{T}}^{W} = \sqrt{2 p_{\mathrm{T}}^{\ell} E_{\mathrm{T}}^{\mathrm{miss}} \left(1 - \cos\Delta\phi\right)} \tag{7.1}$$

  where p_{T}^{ℓ} is the transverse momentum of the lepton and $\Delta\phi$ is the angle between the lepton and the $E_{\mathrm{T}}^{\mathrm{miss}}$.

- $E_{\mathrm{T}}^{\mathrm{miss}}$: The missing transverse energy of the event.

- $E_{\mathrm{T}}^{\mathrm{miss}}$-$\phi$-Centrality: The position in ϕ of the missing transverse energy with respect to the lepton and the visible decay products of the hadronic tau. Defined as

$$E_{\mathrm{T}}^{\mathrm{miss}}\text{-}\phi\text{-Centrality} = \frac{A + B}{\sqrt{A^2 + B^2}}, \tag{7.2}$$

  where

$$A = \frac{\sin\left(\phi_{E_{\mathrm{T}}^{\mathrm{miss}}} - \phi_{\tau}\right)}{\sin\left(\phi_{l} - \phi_{\tau}\right)}, \quad B = \frac{\sin\left(\phi_{l} - \phi_{E_{\mathrm{T}}^{\mathrm{miss}}}\right)}{\sin\left(\phi_{l} - \phi_{\tau}\right)}. \tag{7.3}$$

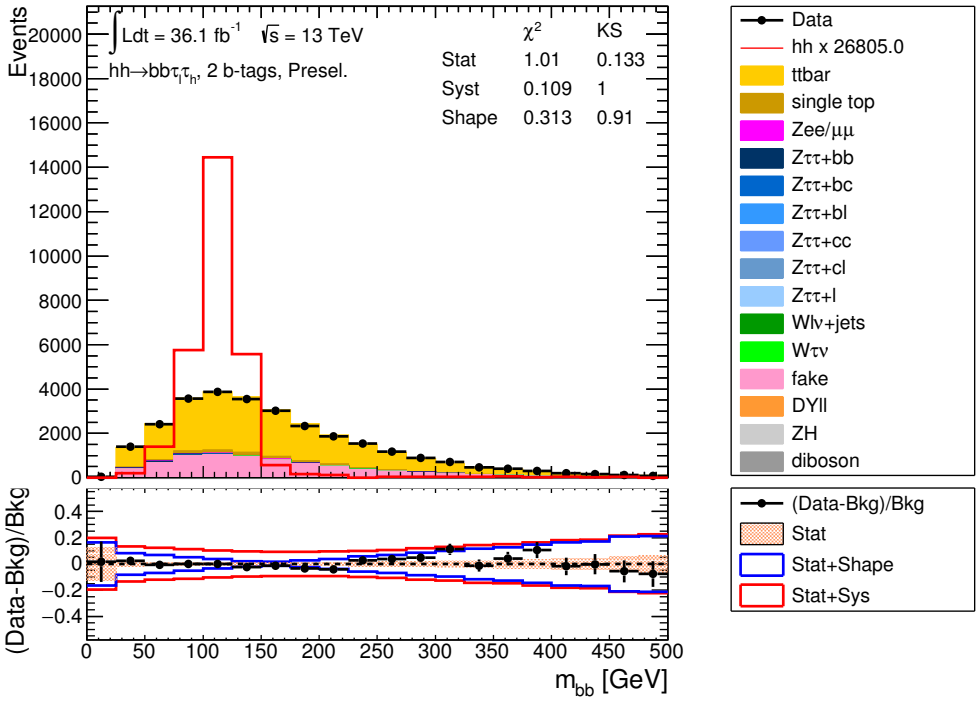
  The $E_{\mathrm{T}}^{\mathrm{miss}}$-$\phi$-Centrality has the values:

  - $\sqrt{2}$ when the $E_{\mathrm{T}}^{\mathrm{miss}}$ is the bisector of the azimuthal, ϕ, angle between the lepton and the tau;
  - $+1$ when the $E_{\mathrm{T}}^{\mathrm{miss}}$ is collinear with either the lepton or the tau;
  - < 1 when the $E_{\mathrm{T}}^{\mathrm{miss}}$ is outside of the angular region ϕ defined by the lepton and tau, -1 when the $E_{\mathrm{T}}^{\mathrm{miss}}$ is anti-collinear with either the lepton or the tau, and decreasing to $-\sqrt{2}$ when the $E_{\mathrm{T}}^{\mathrm{miss}}$ is exactly opposite the lepton and tau.

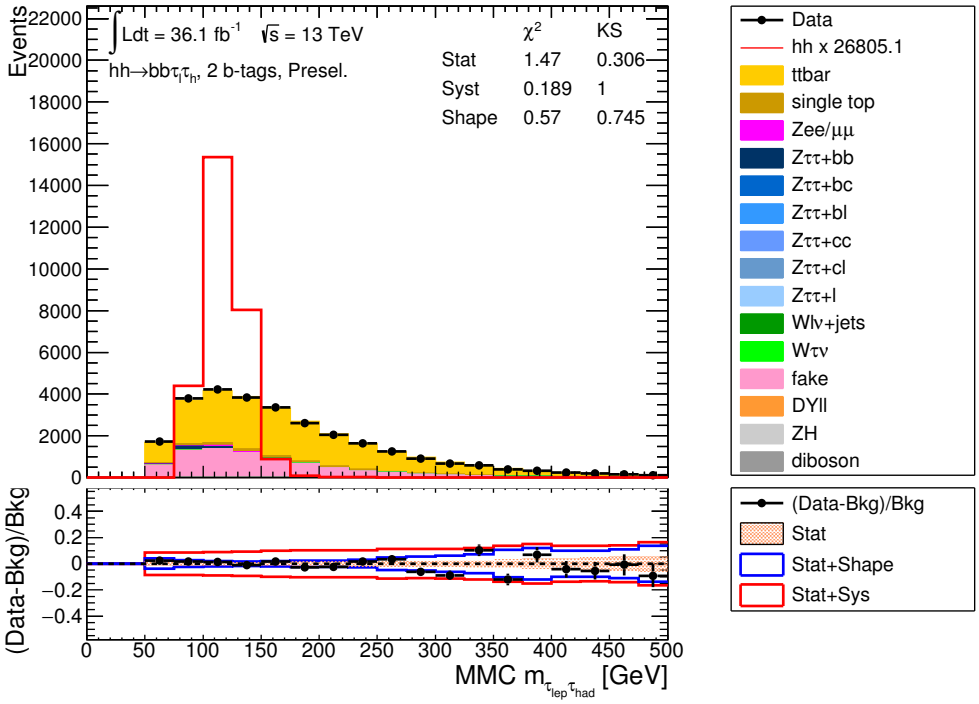
  As the signal has two taus produced from the decay of a Higgs boson, these tend to have higher values of the $E_{\mathrm{T}}^{\mathrm{miss}}$-$\phi$-Centrality as in general the reconstructed $E_{\mathrm{T}}^{\mathrm{miss}}$ lies between the lepton and the tau.

- $p_{\mathrm{T}}^{b_2}$: The transverse moment of the sub-leading b-jet.

Distributions of the input variables are shown in Figures 7.1 to 7.6.

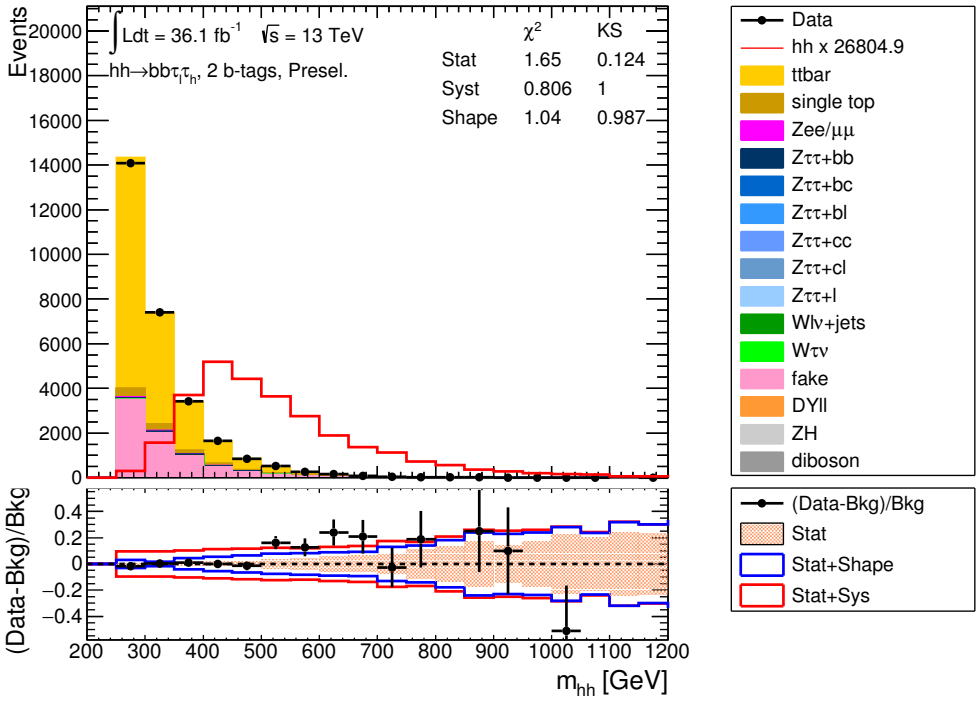


(a)

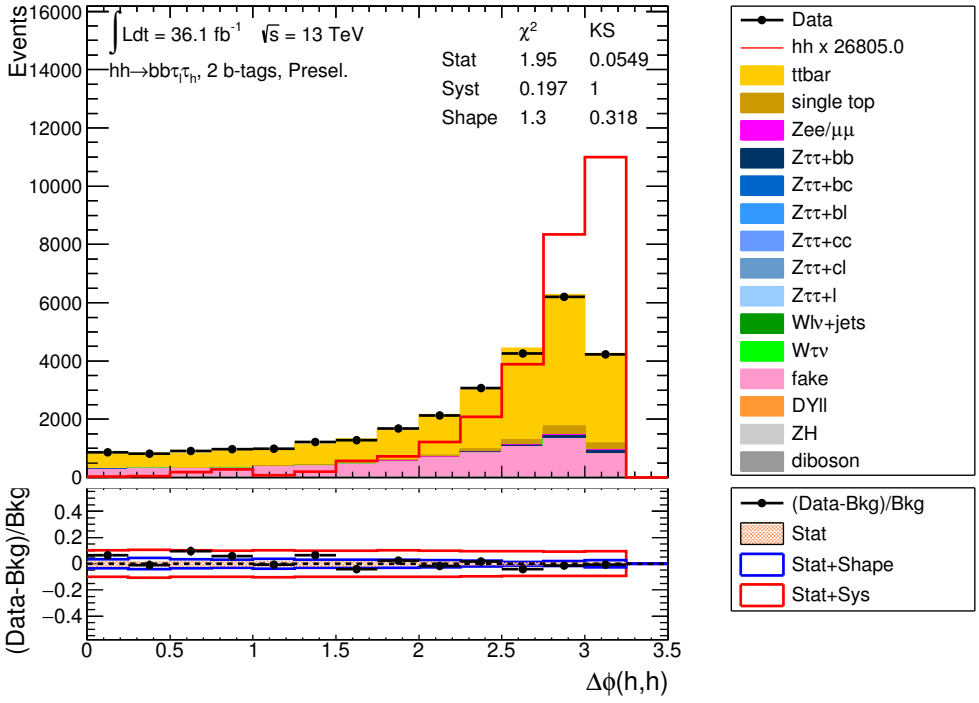


(b)

Figure 7.1: Data-MC comparisons of events at preselection with two $b-$tagged jets of the discriminating variables considered for use in the MVA: (a) m_{bb}, (b) $m_{\tau_\ell \tau_h}^{\mathrm{MMC}}$.

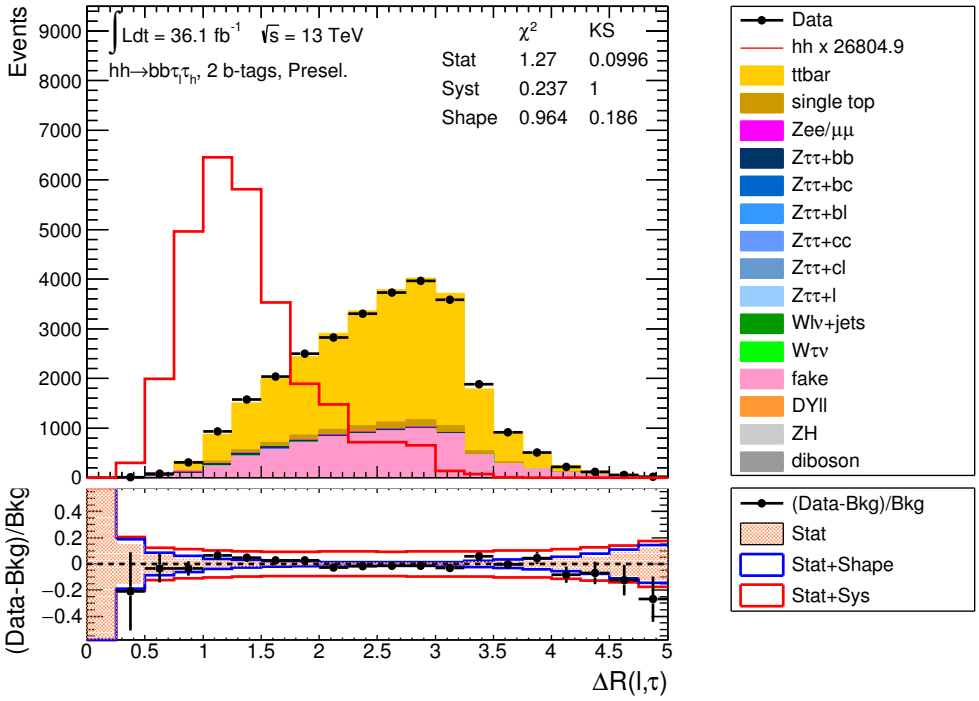


(a)

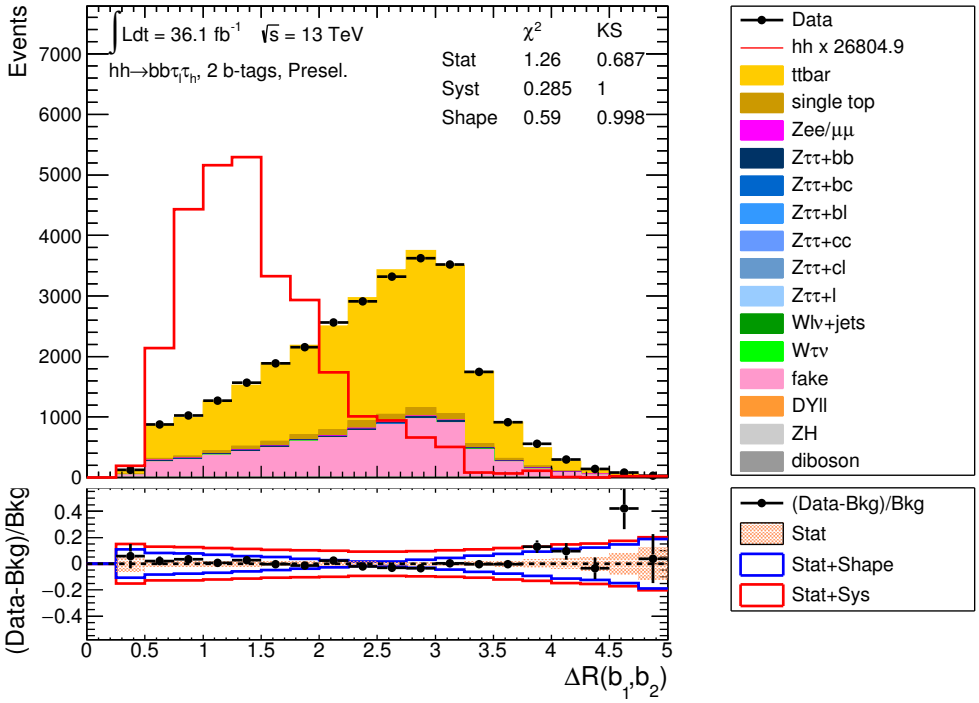


(b)

Figure 7.2: Data-MC comparisons of events at preselection with two $b-$tagged jets of the discriminating variables considered for use in the MVA: (a) m_{hh}, (b) $\Delta\phi\,(h,h)$.

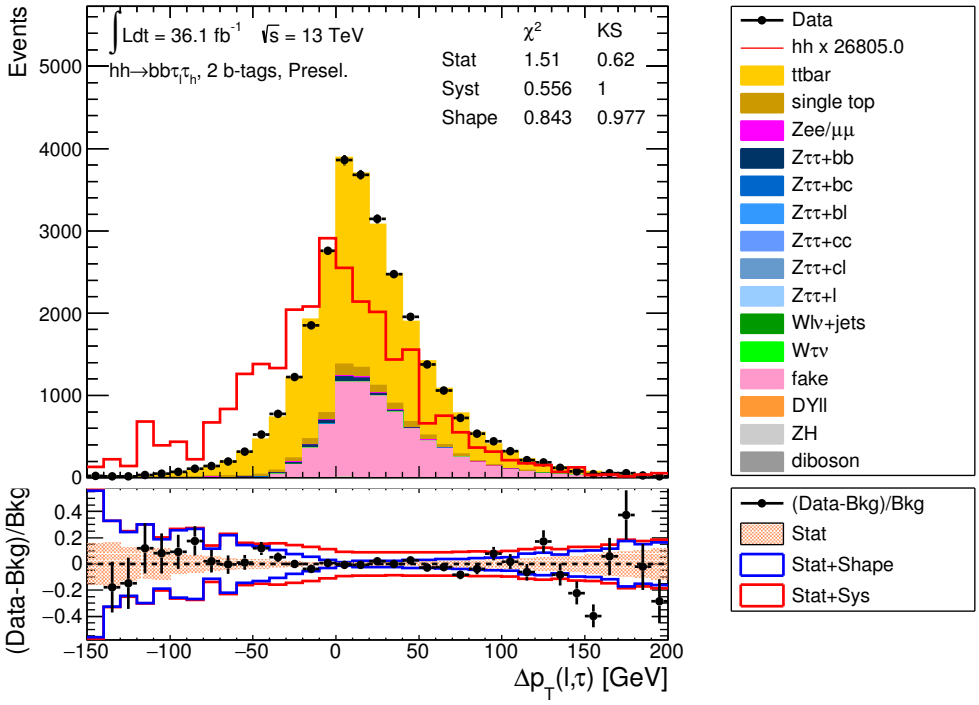


(a)

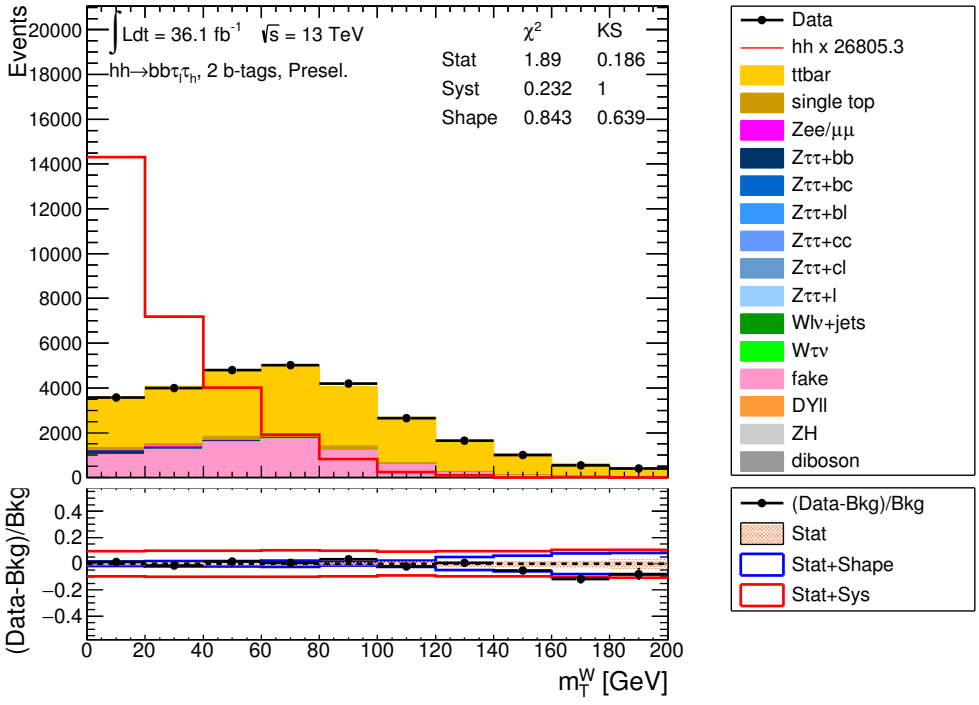


(b)

Figure 7.3: Data-MC comparisons of events at preselection with two $b-$tagged jets of the discriminating variables considered for use in the MVA: (a) $\Delta R\,(l,\tau)$, (b) $\Delta R\,(b_1,b_2)$.

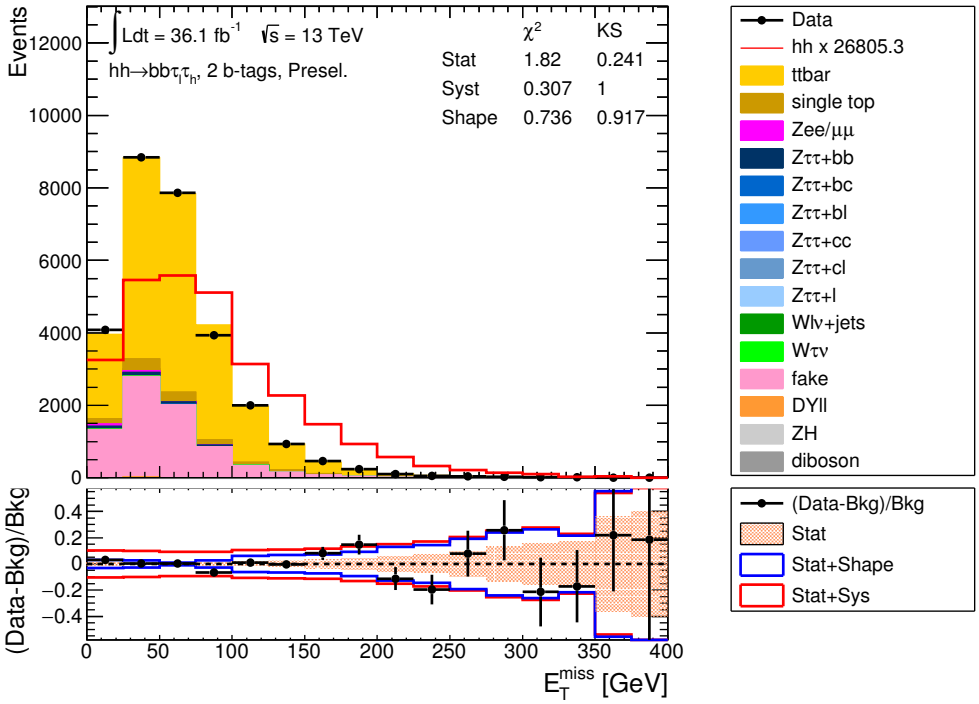


(a)

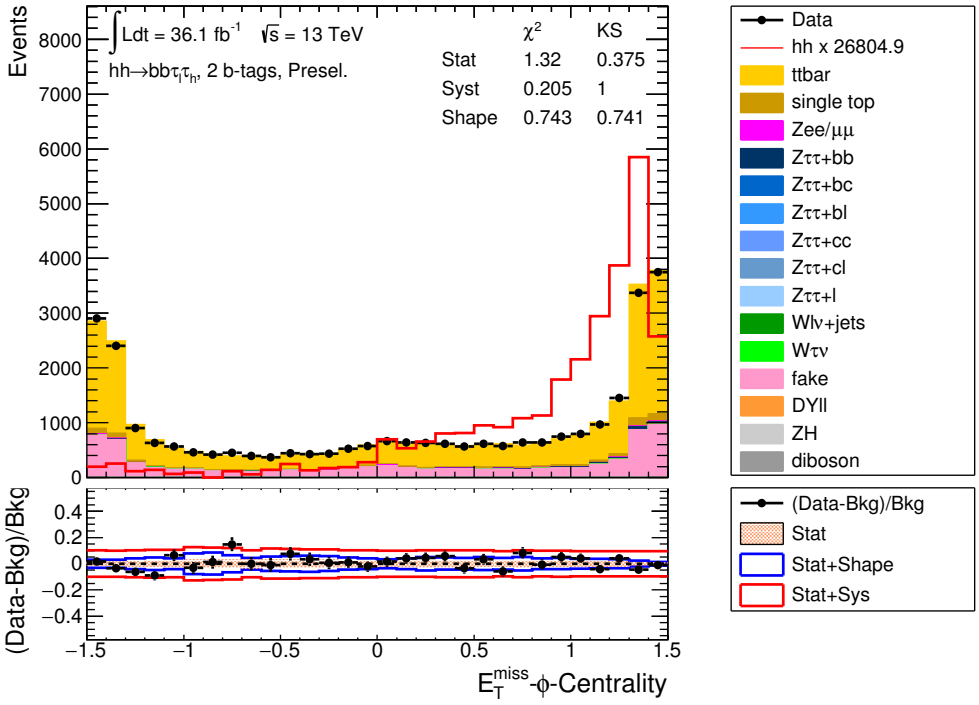


(b)

Figure 7.4: Data-MC comparisons of events at preselection with two $b-$tagged jets of the discriminating variables considered for use in the MVA: (a) $\Delta p_{\mathrm{T}}\,(\ell,\tau)$, (b) m_{T}^{W}.



(a)



(b)

Figure 7.5: Data-MC comparisons of events at preselection with two $b-$tagged jets of the discriminating variables considered for use in the MVA: (a) $E_{\mathrm{T}}^{\mathrm{miss}}$, (b) $E_{\mathrm{T}}^{\mathrm{miss}}$-$\phi$-Centrality.

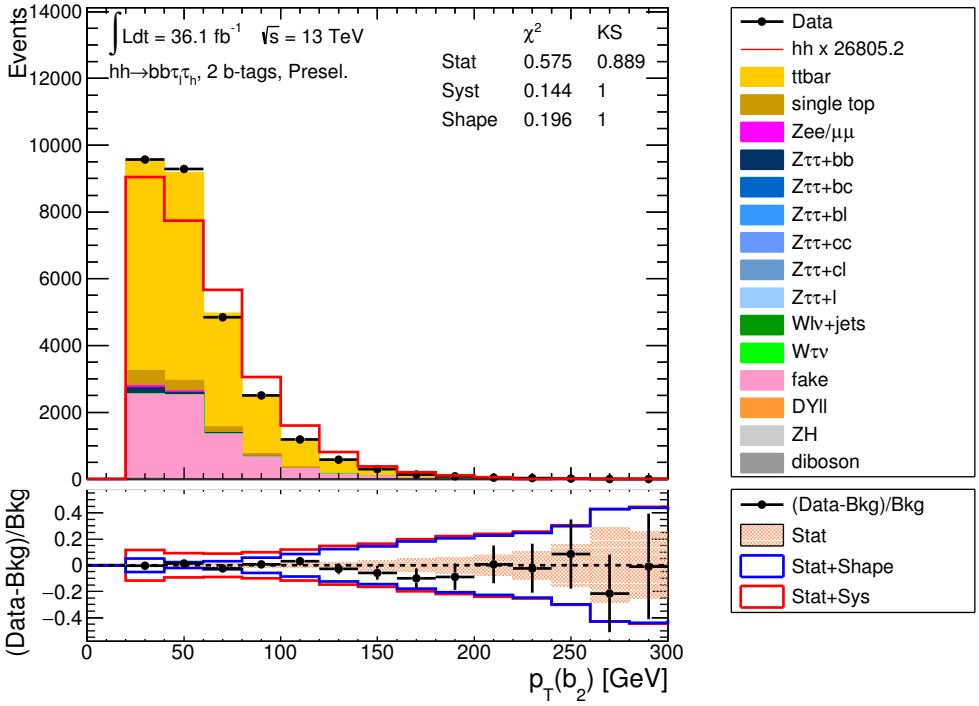


Figure 7.6: Data-MC comparisons of events at preselection with two $b-$tagged jets of the discriminating variables considered for use in the MVA: $p_{\mathrm{T}}^{b_2}$.

7.3.2 Benchmark Trainings

The trainings are performed with MC samples weighted by their predicted cross-sections separately for the 2HDM heavy scalar Higgs, spin-2 RS KK gravitons and the non-resonant SM di-Higgs production. The MVA for the resonance searches are also trained separately for masses between 260 GeV and 1 TeV. These are trained against the dominant $t\bar{t}$ background, where both the real and fake tau components are taken from MC, to maximise rejection of this process. The signal samples are truth b-tagged to improve the statistics for training. For the SLT non-resonant and for all LTT BDTs a subset of the input variables compared to the resonant SLT trainings are used, detailed in Table 7.4, without loss of performance. These are optimised for the low mass region where these channels are most sensitive rather than the full mass range as for the SLT resonant trainings. To reduce overtraining the

Variables	Resonant SLT	Non-Resonant SLT & All LTT
m_{hh}	✓	✓
$m^{\text{MMC}}_{\tau_\ell \tau_h}$	✓	✓
m_{bb}	✓	✓
$\Delta R\,(l, \tau)$	✓	✓
$\Delta R\,(b_1, b_2)$	✓	✓
$E_{\text{T}}^{\text{miss}}$	✓	
$E_{\text{T}}^{\text{miss}}$-$\phi$-Centrality	✓	
m_{T}^{W}	✓	✓
$\Delta\phi\,(h, h)$	✓	
$\Delta p_{\text{T}}\,(\ell, \tau)$	✓	
$p_{\text{T}}^{b_2}$	✓	

Table 7.4: Input variables for analysis BDTs. The non-resonant SLT and LTT trainings use a subset of the resonant SLT variables as they have been optimised for the low mass region where they are more sensitive.

MC datasets are split into two statistically independent sets based on odd and even event numbers, such that they can be used in a 2-fold training procedure (training on odd event numbers, testing on even and vice versa), producing two BDT classifiers for each mass and signal point which can then be applied to the corresponding odd or even events in data. The BDT output distributions for the training and testing samples are compared, separately for signal and background, using the Kolomogorov-Smirnov (KS) test provided in TMVA to check the trainings are not overtrained or seriously affected by statistical fluctuations in the

training data[2]. The configuration of the BDT parameters is also optimised by considering the area under the ROC curves (AUC) and the expected upper limits while monitoring for overtraining[3]. The parameters used for the benchmark trainings are detailed in Table 7.5.

BDT Parameter	Value
BoostType	AdaBoost
AdaBoostBeta	0.15
NTrees	200
MaxDepth	4
MinNodeSize	5%
NCuts	100

Table 7.5: Parameters of the BDTs used for the benchmark trainings.

Signal injection tests were performed, where a signal simulation of a specific mass is inserted to the background simulation and the sensitivity in the limits checked, and demonstrated that BDTs trained on single mass points were insensitive to signals if they were to fall between two MC samples, due to the m_{hh} resolution being as small as 12 GeV for lower resonance masses. In order to solve this problem in addition to the target mass the BDTs were trained on their neighbouring signal mass samples. For example the BDT targeting a 300 GeV mass resonance is trained on the $m_{G,H} = 275,\ 300\,\&\,325$ GeV signal samples, with the cross-section of neighbouring signals normalised to the target signal cross-section. This leads to a slight degradation in the overall performance of the analysis, with a maximum reduction of the upper exclusion limit of 0.10pb (0.15pb) for the 2HDM (RSG) signal, while ensuring sensitivity for signals with masses within the range of the training data.

[2]The values of the KS test from TMVA should be considered with caution as they have a bias from being applied to binned data.

[3]The AUC provides a relatively quick but simple method of evaluating the performance of trainings for comparison. Ideally the eventual FOM for the analysis (in this case the expected upper limits) should ultimately be used, though this can be significantly more computationally intensive.

Chapter 8

Background Estimation

In this chapter summaries of the estimation techniques for the major backgrounds are given. Backgrounds with events containing a real hadronically decaying tau are described in Section 8.1. The data-driven estimation technique used for backgrounds where a jet fakes the hadronic tau, such as $t\bar{t}$ and multi-jet production, are described in Section 8.2. Processes where a Z^0 boson is produced in association with heavy flavour jets are discussed in Section 8.3 and the $Z^0 \to ee$+jets backgrounds are discussed in Section 8.4.

8.1 Backgrounds with a real hadronic tau

MC estimation is used for background events containing true hadronically decaying taus. The normalisation of the $t\bar{t}$ true tau background is allowed to vary in the final fit, with its constraint dominated by the SLT low score region of the BDT distribution. An overall normalisation factor of 1.00 ± 0.12 relative to the nominal MC estimation is obtained from this fit.

8.2 Backgrounds with a jet faking a hadronic tau

A data-driven background estimation method is used for events from processes where a jet is misidentified as the hadronic tau. This involves deriving a tau fake factor (FF) used to weight events from a fake tau 'template' region for each distribution to produce an estimate of these processes. In the 2 b-tag signal region the dominant fake tau contribution comes from $t\bar{t}$, with smaller contributions coming from the multi-jet background, also referred to as QCD. In the control regions for 0 b-tag events the dominant fake contribution comes from W+jets, while in the 1 b-tag region there are approximately equal contributions from W+jets and $t\bar{t}$. Contributions from other background processes are negligible in these regions.

For each of the multi-jet/QCD, $t\bar{t}$ and W+jets backgrounds dedicated control regions

are defined which have enriched fake tau contributions from each process in order to calculate fake-factors for each separately. These control regions are defined with the criteria from preselection with the following additional criteria:

- Multi-jet: The isolation criteria for the electron or muon is inverted ('tight' electrons and 'medium' muons are required to fail their respective 'loose' isolation criteria). Calculated separately for 0 and 1 b-tag events.
- $t\bar{t}$: Calculated in the 2 b-tag region with $m_{\mathrm{T}}^{W} > 40\,\mathrm{GeV}$.
- W + jets: Calculated in the 0 b-tag region with $m_{\mathrm{T}}^{W} > 40\,\mathrm{GeV}$.

The FFs are parametrised in $p_{\mathrm{T}}(\tau_{\mathrm{had}})$ and are calculated separately for 1- and 3-prong tau decays and for each b-tag region, as well as for each process as described above. For the SLT, electron and muon events are combined as the probability for a jet to fake a tau is independent of the flavour of the charged lepton from the leptonic tau decay in the event. For the LTT, FFs are calculated separately for events containing electrons and muons as the associated p_{T} requirements of the light leptons for the trigger are different.
The fake factors are calculated as

$$FF = \frac{N(\tau)}{N(\text{anti-ID-}\tau)}, \tag{8.1}$$

where the anti-ID-τ's are calculated in a fake-tau enriched region, which has the same requirements as the preselection and each of the process specific control regions but with the hadronic tau requirement replaced with the anti-ID-τ requirement as described in Section 6.4.1. Where more than one anti-ID-τ is present in an event, one is selected at random. All derived variables which include the τ_{had} in the analysis, such as the $E_{\mathrm{T}}^{\mathrm{miss}}$, $E_{\mathrm{T}}^{\mathrm{miss}}$-$\phi$-Centrality and $m_{\tau_\ell\tau_h}^{\mathrm{MMC}}$, are then calculated following the same procedure as for the signal region but with the anti-ID-τ replacing the τ_{had}.

The individual FFs for each of the multi-jet, $t\bar{t}$ and W+jets processes are then used to form a combined FF which is applied to an anti-ID-τ 'template' sample, which otherwise has the full signal selection, to correct its normalisation giving an estimate of the total fake tau background contribution in the ID-τ signal region. The combined FF is defined as

$$FF(\text{comb}) = FF(\text{QCD}) \times r_{\text{QCD}} + FF(W/t\bar{t}) \times (1 - r_{\text{QCD}})\,, \tag{8.2}$$

where r_{QCD} is the fraction of fake taus for the multi-jet background in the anti-ID-τ signal region, parametrised as a function of $p_{\mathrm{T}}(\tau_{\mathrm{had}})$:

$$r_{\text{QCD}} = \frac{N(\text{multi-jet}, \text{data})}{N(\text{data}) - N(\text{true}\,\tau_{\mathrm{had}}, \text{MC})}. \tag{8.3}$$

The numerator is given by

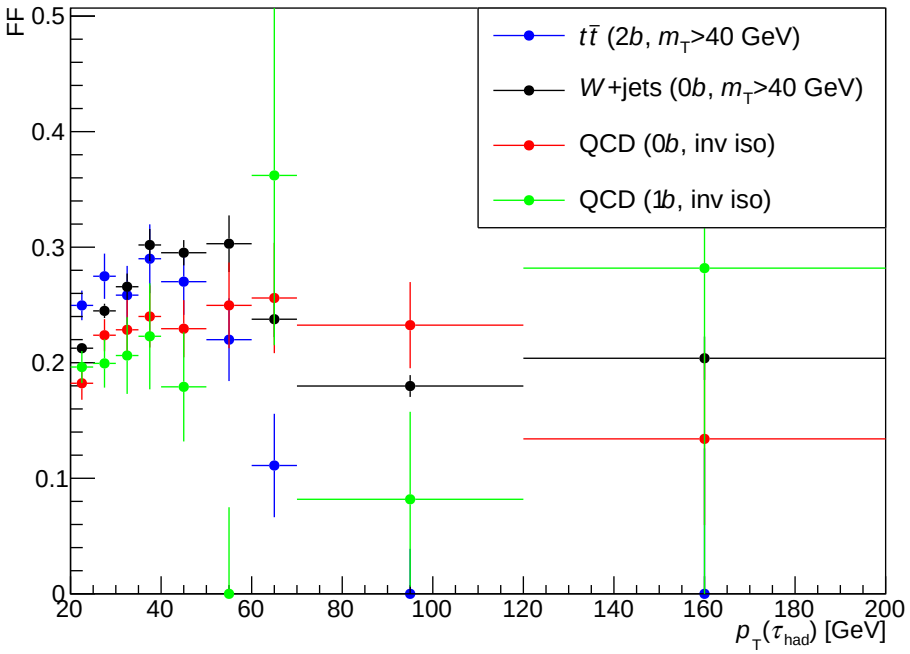
$$N(\text{multi-jet}, \text{data}) = N(\text{data}) - N(\text{true}\,\tau_{\text{had}}, \text{MC}) - N(\text{fake}\,\tau_{\text{had}}, \text{MC}), \tag{8.4}$$

where all predicted MC events from both true and fake taus are subtracted from data in the anti-ID-τ region. The r_{QCD} factor is calculated separately for each b-tag region as well as for electron and muon events as the multi-jet contribution differs for each of these regions. For LTT events the r_{QCD} factor derived in the multi-jet FFs derived in the 1 b-tag regions are used for events with 2 b-tagged jets due to lack of statistics to calculate the FFs in these regions. As the contribution from $t\bar{t}$ (W+jets) is negligible in the 0 (2) b-tag region no attempt is made to account for it when deriving the FFs in this region. The FFs and r_{QCD} for the SLT and LTT channels are shown in Figures 8.1 to 8.7.

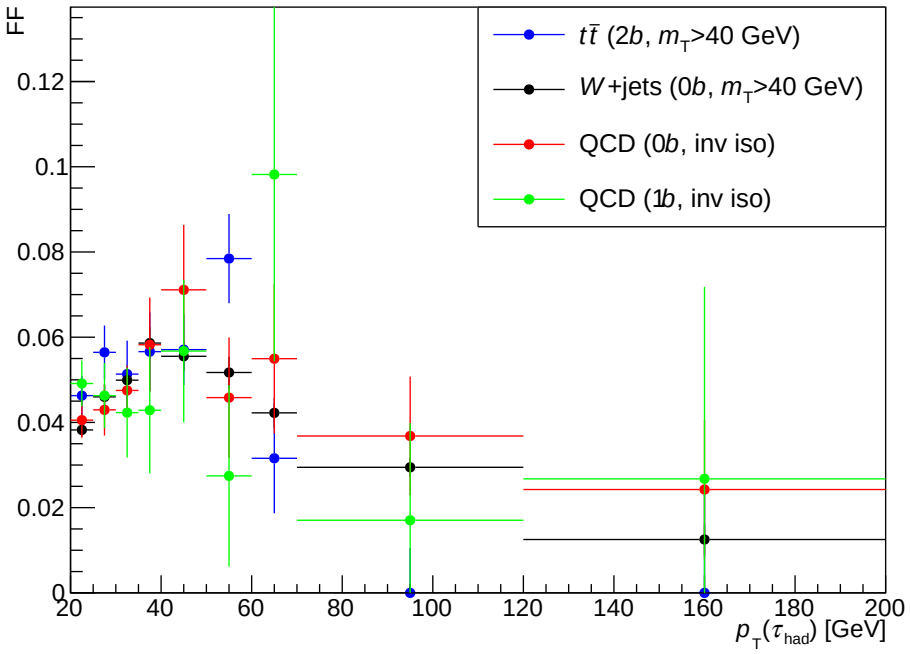
Finally the estimated number of events from fake tau backgrounds are calculated as

$$N^{\text{SR}}_{\text{Est. fake}\,\tau_{\text{had}}} = \left(N^{\text{template}}_{\text{data}} - N^{\text{template}}_{\text{true}\,\tau_{\text{had}},\text{MC}}\right) \times FF_{\text{comb}}, \tag{8.5}$$

where all simulated MC true tau events, and the small fake tau component from the negligible backgrounds, are subtracted from the data in the 'template' region for each distribution.

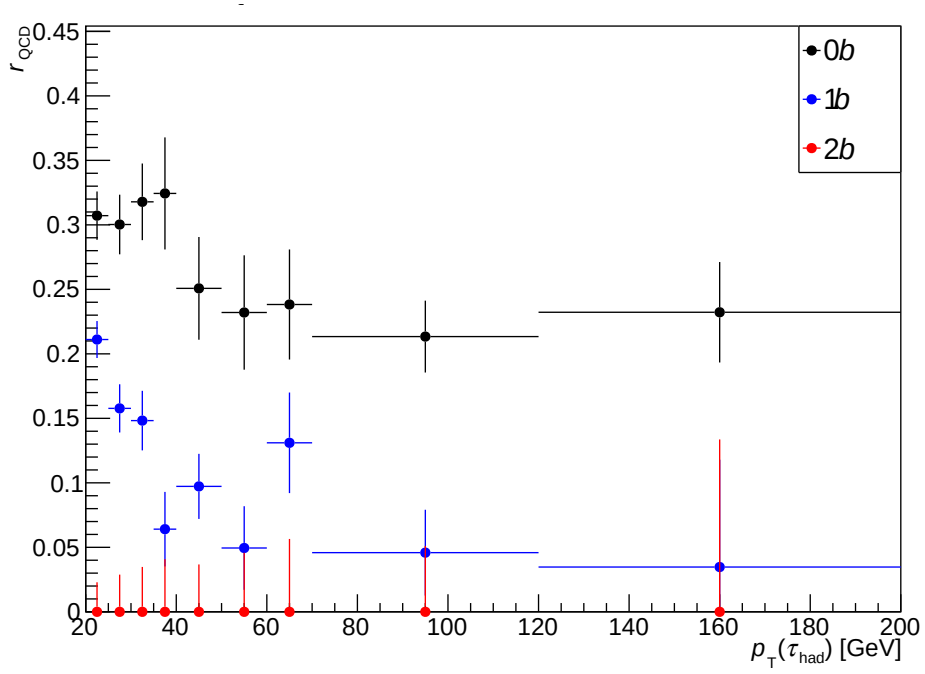


(a)

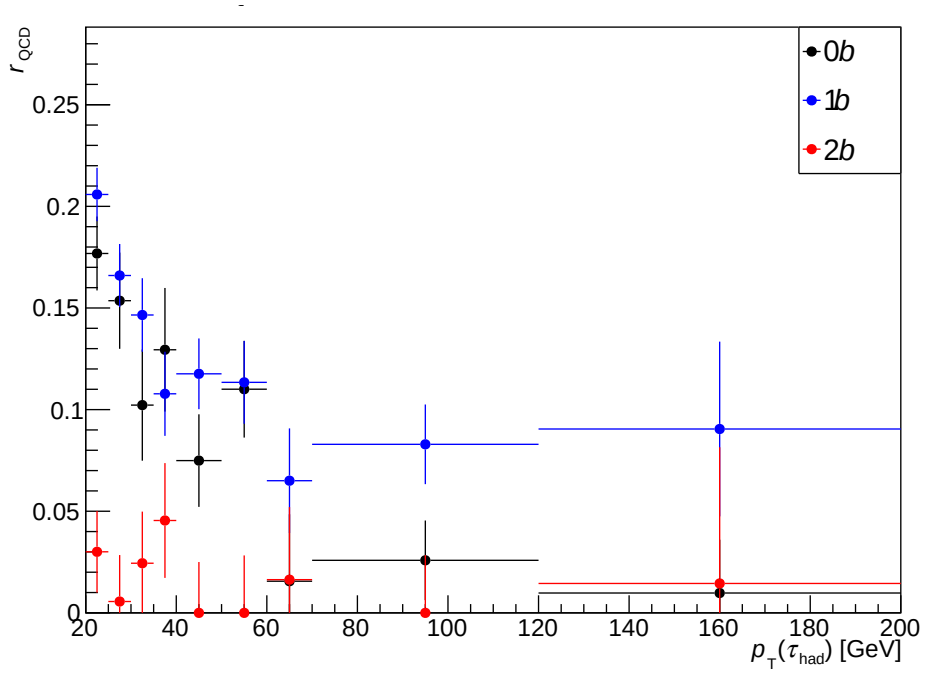


(b)

Figure 8.1: Fake factors (FF) for 1-prong (a) and 3-prong (b) taus in the SLT channel. The fake factors are calculated in separate control regions for the $t\bar{t}$, W+jets and QCD processes and are inclusive in leptons.

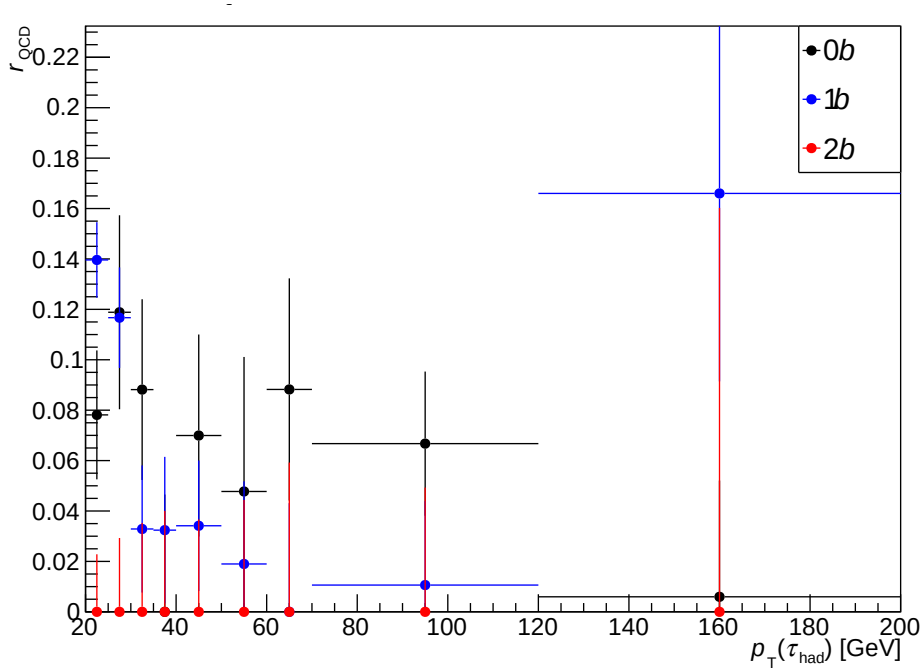


(a)

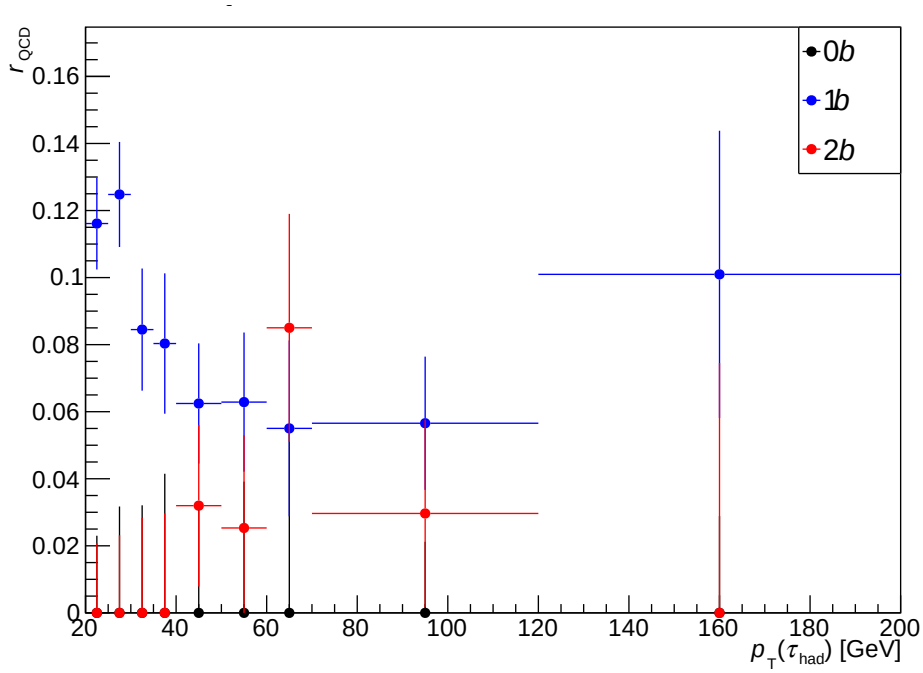


(b)

Figure 8.2: r_{QCD} for 1-prong (a) and 3-prong (b) taus in the $e\tau_{\mathrm{had}}$ SLT channel. Calculated separately for the zero, one and two b-tag categories.

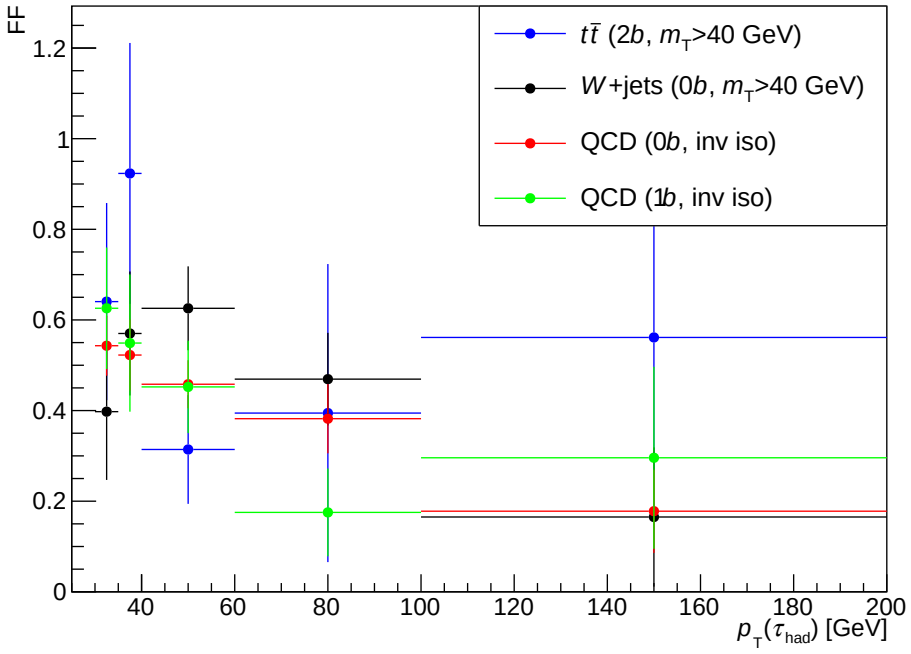


(a)

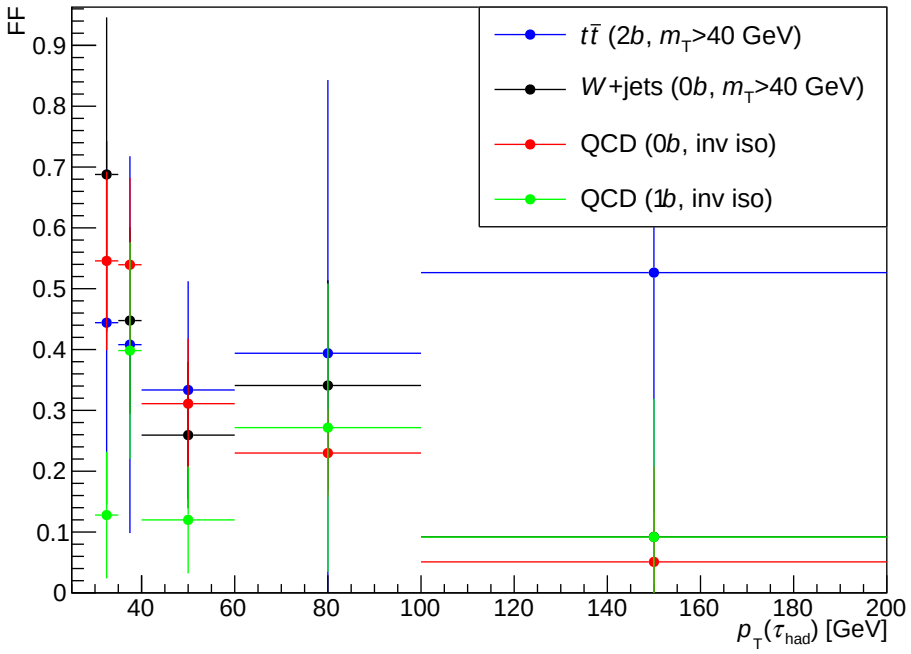


(b)

Figure 8.3: r_{QCD} for 1-prong (a) and 3-prong (b) taus in the $\mu\tau_{\mathrm{had}}$ SLT channel. Calculated separately for the zero, one and two b-tag categories.

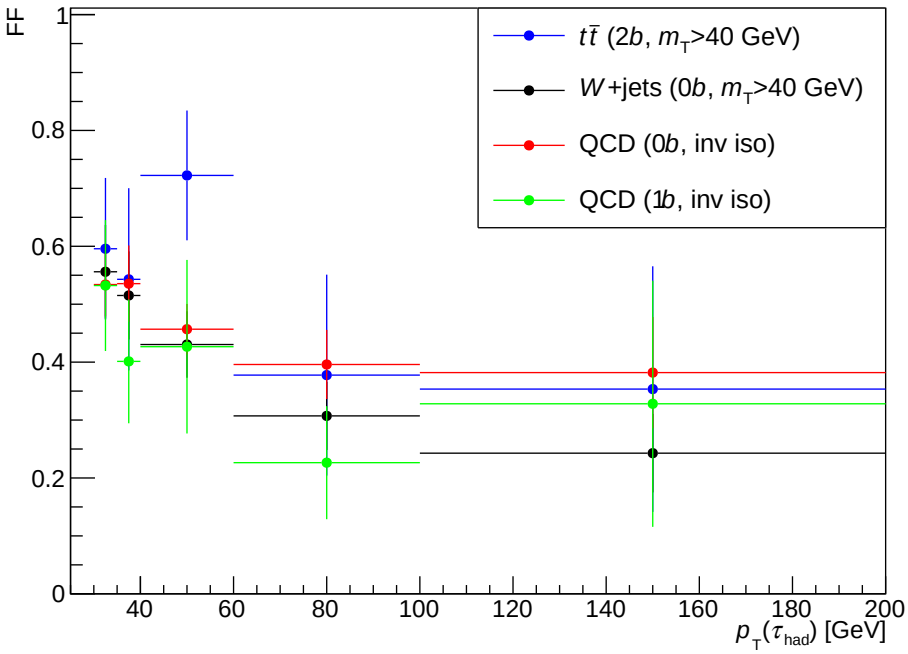


(a)

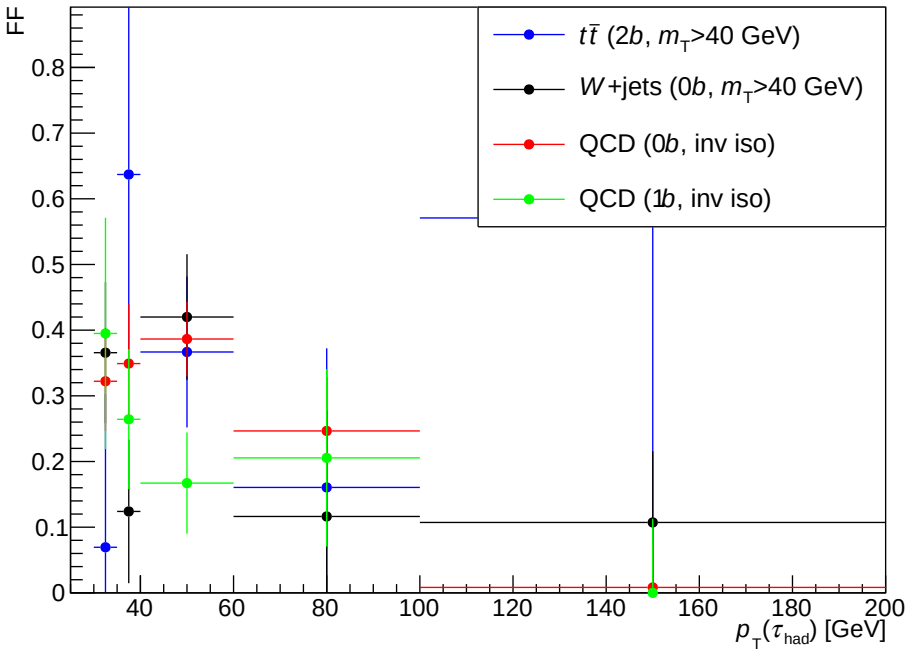


(b)

Figure 8.4: Fake factors (FF) for 1-prong (a) and 3-prong (b) taus in the $e\tau_{\text{had}}$ LTT channel. The fake factors are calculated in separate control regions for the $t\bar{t}$, W+jets and QCD processes.

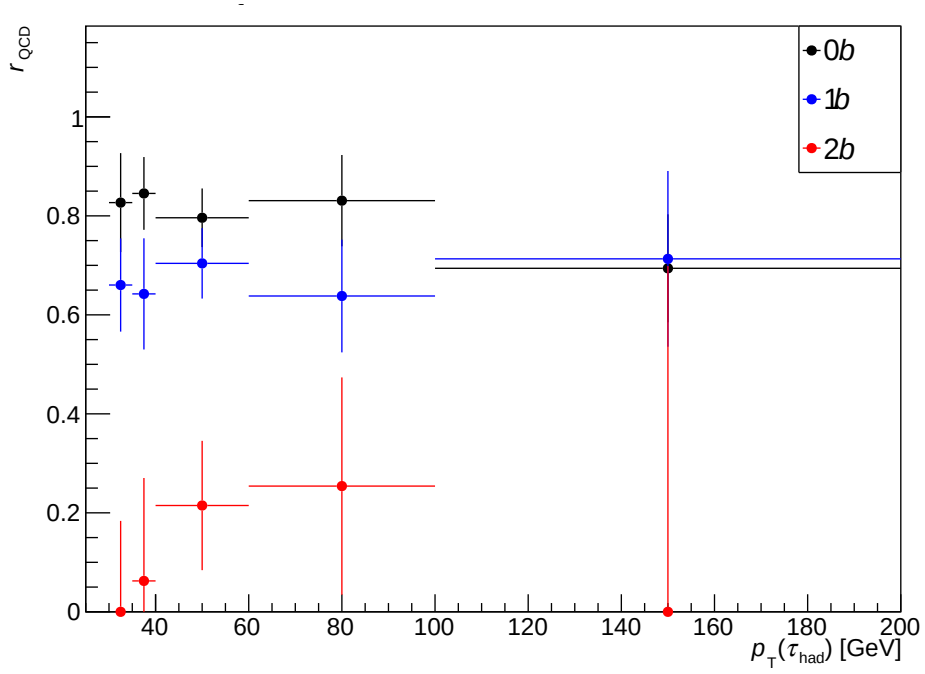


(a)

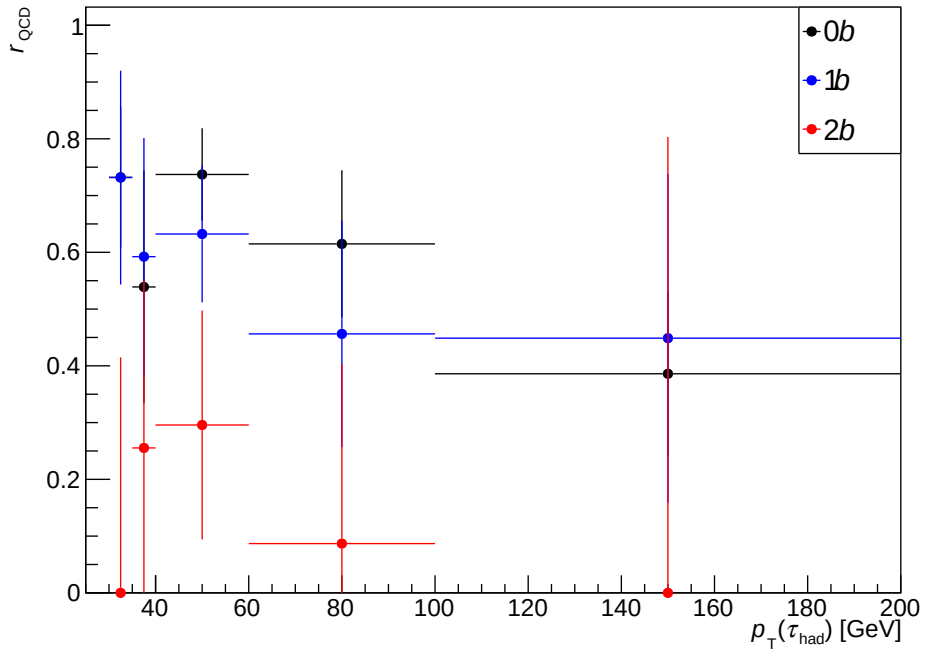


(b)

Figure 8.5: Fake factors (FF) for 1-prong (a) and 3-prong (b) taus in the $\mu\tau_{\text{had}}$ LTT channel. The fake factors are calculated in separate control regions for the $t\bar{t}$, W+jets and QCD processes.

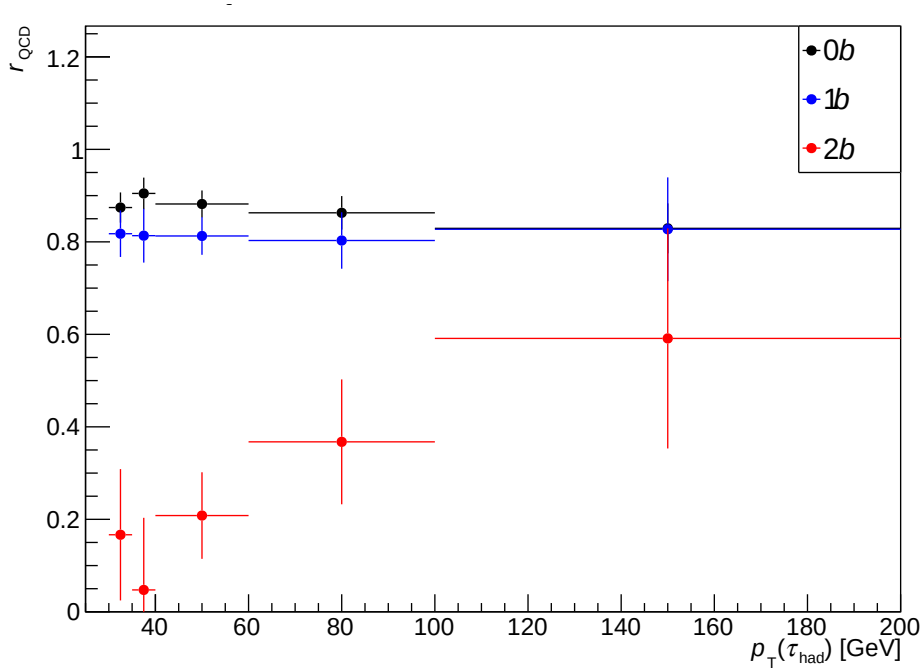


(a)

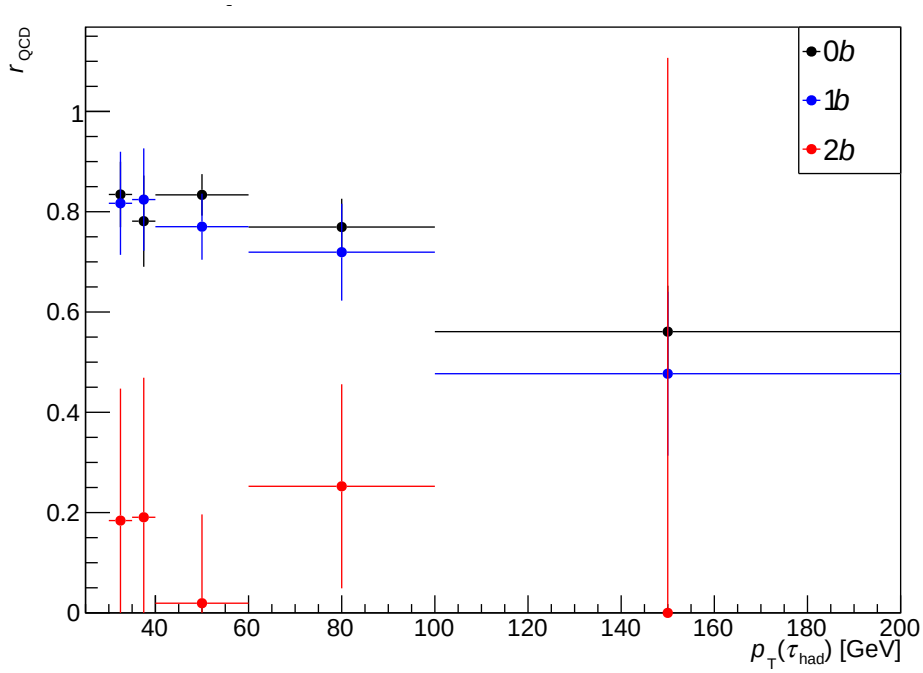


(b)

Figure 8.6: r_{QCD} for 1-prong (a) and 3-prong (b) taus in the $e\tau_{\mathrm{had}}$ LTT channel. Calculated separately for the zero, one and two b-tag categories.



(a)



(b)

Figure 8.7: r_{QCD} for 1-prong (a) and 3-prong (b) taus in the $\mu\tau_{\mathrm{had}}$ LTT channel. Calculated separately for the zero, one and two b-tag categories.

8.3 Z + heavy flavour jets

For the processes where a Z^0 boson produced in association with heavy flavour (b, c) jets (Z+HF), it is known that the MC simulation does not describe the data well and as such a control region (CR) is used to normalise the MC to the data. As these processes should be independent of the decay mode of the Z^0 boson, for the CR events where $Z^0 \to \mu^+\mu^-$+ two heavy flavour jets (bb, bc, cc) are selected from data. These events are required to pass the single muon triggers (as described in Section 7.1), contain exactly two muons with $p_{\mathrm{T}} > 27\,\mathrm{GeV}$ and a di-muon invariant mass between $81\,\mathrm{GeV} < m_{\mu\mu} < 101\,\mathrm{GeV}$, and at least two b-tagged jets (fulfilling the b-tagging requirements in Section 6.1.1) with $p_{\mathrm{T}} > 45(20)\,\mathrm{GeV}$ for the leading (sub-leading) b-jet. The di-b-jet mass is required to be $m_{bb} < 80\,\mathrm{GeV}$ and $m_{bb} > 140\,\mathrm{GeV}$ in order to veto events from the SM $Vh(h \to bb)$ production.

This CR is included in the final fit as a single bin histogram and is used to determine the Z+HF normalisation. As with the true τ $t\bar{t}$ background, the Z+HF normalisation is allowed to vary in the final fit; the assumptions are made that the backgrounds from $Z \to \tau\tau$ and $Z \to \mu\mu$ are correlated in the two regions, and the $t\bar{t}$ background is correlated across the regions. The statistical uncertainties on the the data and the MC are taken into account, as are the contributions from the SM $Zh, h \to bb$ background. For each signal region fit separate normalisation factors are derived and applied to all $Z^0 \to \ell\ell(e, \mu, \tau) + jj(bb, bc, cc)$ processes. The background-only fit obtains a normalisation factor of 1.45 ± 0.15 relative to the nominal MC expectation, where the uncertainty quoted is fully profiled as it is obtained from the limit fit. The background-only post-fit yields of the different background processes for this CR are given in Table 8.1.

Samples	Post-fit yield
$Z^0 \to \ell\ell + (bb, bc, cc)$	8540 ± 450
Top quark	4030 ± 420
Other ($W^{\pm}$ & $Z^0 + (bl, cl, ll)$ & VV)	320 ± 120
Total SM Background	12890 ± 110
Data	12897

Table 8.1: Post-fit event yields for the $Z^0 \to \mu^+\mu^- + (bb, bc, cc)$ control region for a background-only fit. The 'Top quark' category includes contributions from $t\bar{t}$ and single top production. The 'Other' category includes contribution from the $W^{\pm}$+jets, Z^0+light flavour jets and di-boson processes.

8.4 $Z \to ee$ + jets

A CR is utilised to check the normalisation of the $Z \to ee$+jets backgrounds. It is defined by applying the standard event preselection and additionally requires that the $m_{\mathrm{T}}^{W} < 40\,\mathrm{GeV}$, the $E_{\mathrm{T}}^{\mathrm{miss}}$-$\phi$-Centrality$> 0$, the visible di-tau mass is $81\,\mathrm{GeV} < m_{\tau\tau}^{\mathrm{vis}} < 101\,\mathrm{GeV}$ and the event contains either zero or one b-tagged jet. The normalisation factor for the Z+ heavy flavour jets background (described in Section 8.3) is applied to the $Z \to ee + bb, bc, cc$ contributions. For the zero b-tag region the normalisation factor is found to be 0.81 ± 0.06, while in the one b-tag CR the normalisation factor is found to be consistent with unity. No normalisation factor is applied to the central value of the $Z \to ee + b$-jets background and instead an uncertainty is applied to cover the difference between unity and the zero b-tag normalisation factor in order to account for any mis-modelling in the rate of electrons faking the decay of hadronic taus.

Chapter 9

Systematic Uncertainties

There are many sources from which systematic uncertainties can arise in the analysis. These can impact the analysis in two ways; shape uncertainties that affect the shapes of the distributions of variables used such as the inputs to the MVA, and normalisation uncertainties which affect the expected yields of the signal and background. Both types ultimately affect the number of events in the final distributions used in the final fit to extract results. Each source of uncertainty is propagated into the final result as a nuisance parameter term in the likelihood function (described in Chapter 10), with the functional dependence of the uncertainties of both the expected number of signal and background events also taken into account in the likelihood. The uncertainties are grouped into three categories: detector uncertainties, background modelling uncertainties and theoretical uncertainties, as described in the following with the main sources of uncertainty summarised in Table 9.1. The dominant systematic uncertainties arise from the modelling of the $t\bar{t}$ background, uncertainties related to the reconstruction and identification of hadronic taus and those related to the b-tagging methods.

9.1 Detector

Systematic uncertainties related to the detector simulation are evaluated for the signal samples and for the backgrounds using the MC simulation.

9.1.1 Luminosity

The scale of the luminosity is calibrated using x-y beam-separation scans (van-der-Meer scans) [89], performed in August 2015 and May 2016. The uncertainty on the combined 2015+2016 dataset is measured to be 2.1%.

9.1.2 Jet energy scale and resolution

The uncertainties related to jets are associated with the jet energy scale (JES), jet energy resolution (JER) and the efficiency to pass the jet vertex tagger (JVT) algorithm [77]. There are a number of systematic uncertainties related to the JES calibration, which is used to correct the MC to account for effects from sources such as the jet reconstruction algorithm, jet fragmentation, calorimeter response and pile-up interactions. In-situ measurements and comparisons between simulation and data are used to determine the calibrations and corresponding uncertainties [90]. The JES uncertainties are parametrised in bins of p_{T} and η and are grouped into three reduced sets of uncertainties, while preserving their correlations. The JES calibration uncertainty is largest for low p_{T} jets starting at 4.5% for 20 GeV jets and decreasing to 1% for 200 GeV jets, rising to 2% at 2 TeV mainly due to statistical uncertainties. The JER uncertainty was measured in Run 1 [91] by comparing the MC simulation to data, parametrised in p_{T} and η, which were found to agree within 10%. The total uncertainty on this measurement was between 10% and 20%. The JVT algorithm is applied to reduce the jets from pile-up events and the threshold on the output of this is varied as an additional uncertainty. Both the JER and JVT uncertainties were updated for Run 2 with additional systematic uncertainties to account for the different conditions [92].

9.1.3 Jet flavour tagging

Scale factors are applied to account for the differences in flavour-tagging efficiency between data and simulation. These are measured separately for b-, c- and light-flavour jets [78]. The correction factors are parametrised in jet p_{T} for b- and c-jets and in jet p_{T} and η for the light-jets. These have a number of sources of uncertainty, which are decomposed into uncorrelated sets applied independently with three uncertainties for b-jets, four uncertainties for c-jets and five for light-flavour jets. The correction factors for b-jets are derived using $t\bar{t}$ events from data with the measurements resulting in uncertainties between 2% and 4% for jets with $20\,\mathrm{GeV} < p_{\mathrm{T}} < 200\,\mathrm{GeV}$, rising to 12% for jet p_{T} up to 300 GeV. Uncertainties are also applied to account for the extrapolation of the b-jet efficiencies for $p_{\mathrm{T}} > 300\,\mathrm{GeV}$. Multi-jet events and $W+c$-jet events from data are used to derived the calibrations and their uncertainties on the mis-identification of c- and light-jets as b-jets, with the uncertainties on the efficiency being 5%–13% and 20%–50% respectively depending on the b-tagging working point.

9.1.4 Electrons

Uncertainties arising from the electron trigger, identification and isolation requirements are taken into account. These are derived using the tag-and-probe method where a well

reconstructed and identified electron is used to select the event and another electron with minimal selection is used to evaluate the efficiencies. This is performed on samples of $Z^0 \to ee$ and $J/\psi \to ee$ events using all $3.2\,\text{fb}^{-1}$ of 2015 data [93]. Parameterised as functions of the electron transverse energy, E_T, and η, the variations range from below 1% up to 5%.

9.1.5 Muons

Scale factors and smearing corrections are applied to simulated muons to account for discrepancies between data and MC coming from the muon momentum scale and energy resolution. Calculated as a function of muon p_T and η, samples from the full 2015 dataset are used with $Z^0 \to \mu\mu$ used for the high p_T region ($p_\text{T} > 20\,\text{GeV}$) and $J/\psi \to \mu\mu$ for the low p_T region ($p_\text{T} < 20\,\text{GeV}$) [82] leading to variations of the muon p_T between 1.7% and 2.9%. Uncertainties on the identification, isolation and trigger efficiencies are also considered, utilising the same $Z^0 \to \mu\mu$ and $J/\psi \to \mu\mu$ samples. The variation due to these uncertainties are found to be between 1% and 7%.

9.1.6 Taus

The performance of the online and offline hadronic tau reconstruction and identification were measured using the 2015 $3.2\,\text{fb}^{-1}$ dataset [85]. The energy calibration and the efficiency of the identification algorithms are measured using $Z^0 \to \tau\tau$ tag-and-probe measurements; the uncertainty on the offline tau identification are 5% (6%) for one (three) prong taus with tau p_T greater than 20 GeV and inclusive in η, and the uncertainty for the online identification is between 3% and 20% depending on tau p_T. The probability of mis-identifying an electron as a hadronic tau is $< 2.5\%$ with a corresponding uncertainty of 3%–14% and the uncertainty on the tau energy scale is measured to be 2% to 3%.

9.1.7 E_T^miss

The uncertainties on the energy resolution and scale of the jets, electrons, muons and taus (described above) are propagated to the calculation of the E_T^miss. Additional dedicated variations are also included to account for the uncertainties on the scale, resolution and reconstruction efficiency of tracks that are not associated to any of the reconstructed objects, as well as any mis-modelling of the underlying event.

9.2 Background Modelling

9.2.1 Uncertainties on top quark processes

As the $t\bar{t}$ background process containing jets faking taus is estimated from data, generator level uncertainties are only applied to the $t\bar{t}$ background component containing real hadronic taus and include those associated with the shower radiation and hadronisation modes. These systematic effects are estimated by varying the factorisation and renormalisation scales of Powheg+Pythia 6 up and down by a factor of 2 and the fragmentation model is compared to the simulation from Powheg+Herwig. A comparison between a sample generated by aMC@NLO and showered with Herwig++ to a sample generated by Powheg and showered with Herwig++ is used to account for additional uncertainties arising from the hard scatter generation. The differences in shape between the nominal and these variations are parameterised as a function of m_{bb} and p_{T}^{bb}.

9.2.2 Uncertainties on backgrounds where a jet fakes a hadronic tau

As the composition of fake-taus differs between the signal region and the CR a systematic uncertainty is assigned to account for this. A CR which requires the tau and lepton have the same-sign (SS) charge as well as the event containing 2 b-tags is used. The m_{hh} shows the greatest discrepancy between data and MC of all the BDT input variables for this region and has the largest variation of fake-tau composition across the distribution, so is chosen to parametrise this uncertainty. The ratio of the data to MC of the m_{hh} distribution is fitted to obtain a variation on the output BDT score, with the fit being symmetrised to provide both the up and down variations.

The effect of the detector related uncertainties (described in Section 9.1) on the fake-tau estimation is determined by applying the up and down variation to the true-tau $t\bar{t}$ contribution and the estimation rederived. The difference between the resulting distributions and the nominal fake-tau background estimation is used to estimate the shape uncertainty and is parametrised in the tau p_{T}, shown in Figure 9.1, using the same binning scheme as the fake-factors. The same procedure is repeated varying the true-tau $t\bar{t}$ component up and down by the theoretical uncertainties on the $t\bar{t}$ modelling (described in Section 9.2.1) as well as for the NLO cross-section uncertainty on $t\bar{t}$ production of 6%. All other background processes in the control regions used to derived the background estimation, which have a negligible contribution, are varied up and down by 50%. The fake-factors are also allowed to vary up and down by their statistical uncertainties as an additional uncertainty.
To account for the uncertainty in the fake tau component of the $t\bar{t}$ background separate fake-factors are calculated from MC using the nominal selection and a high m_{T}^{W} region ($m_{\mathrm{T}}^{W} > 40\,\mathrm{GeV}$). The difference in each bin between these and the nominal fake-factors

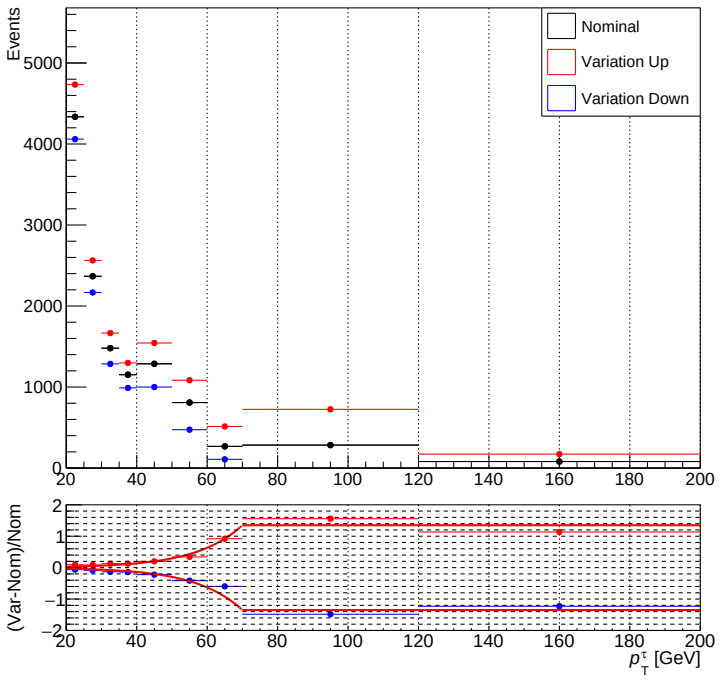


Figure 9.1: Effect on the fake-tau background estimation of the up (red points) and down (blue points) systematic variations arising from detector related uncertainties compared to the nominal (black points). The difference from the nominal distribution is used to parameterise the shape uncertainty as a function of tau p_{T} shown by the red line.

derived from data are used as up and down variations with the fake tau estimate recalculated for each case. The shape of this uncertainty was found to be flat and so is included as a normalisation uncertainty.

All the above uncertainties are included as shape and normalisation uncertainties in the fit, unless stated otherwise. For each of the systematic variations $\mathrm{r_{QCD}}$ is recalculated when deriving the fake tau estimation. To estimate the uncertainty on $\mathrm{r_{QCD}}$ itself a similar process as described above for the fake factors is performed. These are included as shape and normalisation uncertainties in the final fit. No additional uncertainty is assigned to account for backgrounds where leptons fake taus, estimated entirely from simulation, since these processes make up less than 5% of the total background and even an overly conservative uncertainty would have a negligible contribution in the final result.

9.2.2.1 Uncertainties on the $W^{\pm}$, Z^0, Higgs and diboson processes

For all other background processes estimated from MC the theoretical uncertainties on the cross-sections are considered. The fit to the $Z^0 \to \mu\mu+$ heavy flavour jets CR, described in Section 8.3, is used to derive the uncertainty on the Z^0+ heavy flavour jets (bb, bc, cc)

processes and is found to be 10.3% for a background-only fit when including all sources of systematic uncertainty. The uncertainty for the Z^0+ light flavour jets (cl, ll) and diboson processes are 5% and 6% respectively when combining the scale, α_s and PDF variation uncertainties in quadrature, following the procedure in Reference [94]. For backgrounds from $W^{\pm}$+ jets production an uncertainty of 20% is assigned to cover the uncertainties associated with the $W^{\pm}$+ heavy flavour jets processes.

Uncertainties due to the choice of PDF set are evaluated using event weights included in the Sherpa 2.2.1 samples, comprised of 100 variations on the nominal NNPDF3.0 PDF set as well as central values from two alternative PDF sets, MMHT2014nnlo68cl and CT14nnlo. The standard deviation of the 101 NNPDF3.0 PDFs is taken as the estimate of the intra-PDF uncertainty and the envelope of the differences between the two alternative PDF sets and the nominal set are included as additional uncertainties. The uncertainties relating to the choice of the renormalisation and factorisation scales on the Z^0+ jets background modelling are evaluated by varying the scales up and down by a factor of 2 either together or independently using the event weights included in the samples. The total uncertainty related to choice of scale is then taken as the envelope of these variations. Alternative samples generated using MG5_aMC@NLO v2.2.2 at Leading Order and interfaced to the Pythia 8.186 parton shower model are compared to the nominal Z^0+ jets samples. These use the A14 tune together with the NNPDF2.3LO PDF set and EvtGen v1.2.0 is used for the properties of the bottom and charm hadron decays. For all of the above sources of systematic uncertainties the m_{bb} and p_{T}^{bb} distributions are used to parameterise the discrepancy between the nominal and variations. These are then propagated through the analysis as a shape uncertainty on the BDT.

As described in Section 8.4, the normalisation of the $Z^0 \to ee$ background in the 1 b-tag $Z^0 \to ee$ is found to be consistent with unity. As such no correction is applied to the normalisation for the 2 b-tag signal region, and instead the difference between the normalisation correction scale factor derived in the 0 b-tag $Z^0 \to ee$ control region and 1.0, 19% is applied to account for any mis-modelling between data and MC of the rate for electrons to fake hadronic taus.

For the background component from associated Zh production a normalisation uncertainty of 50% is applied.

9.3 Theory

The theoretical uncertainties on the 2HDM, Graviton and non-resonant signal processes are evaluated by varying the PDF sets used, the renormalisation scale that affects both α_s

and the PDFs, and the uncertainties related to the PDF sets themselves. For the resonant signal models, the uncertainty due to the choice of PDF set on the signal cross-section is evaluated by comparing the nominal NNPDF23 PDF set to the PDF4LHC set. The 100 eigenvectors of NNPDF23 set are also varied and compared to the nominal. The uncertainty from these sources are found to range from 4–15% (7%) for the 2HDM (Graviton) for the PDF set variation and between 2% and 4% for the eigenvector variations for both signal processes. Higher order corrections to the signal production cross-sections are evaluated by varying the QCD renormalisation and factorisation scales by a factor of 2. The variation is found to be $\sim 15\%$ for the 2HDM signals and range from 3% to 6% for the Graviton signals.

The uncertainty on the signal acceptance due to variations on the assumptions about the underlying event (UE), initial state radiation (ISR) and final state radiation (FSR) are calculated by generating the signal samples with different Pythia 8 tunes. In order to evaluate each variation, the BDT input variables are constructed from truth level variables and are used to evaluate the effect on the normalisation and shape of the BDT output distribution. The total normalisation difference is found to be between 2% and 7%, depending on the signal model.

For the non-resonant Higgs pair production the uncertainty on the cross-section is taken to be $33.49^{+1.44}_{-2.00}(\text{scale}) \pm 1.67(\text{theory}) \pm 0.77(\alpha_s) \pm 0.70(\text{PDF})$ fb [6, 95].

Systematic Uncertainty	Type	Components
Luminosity	N	1
Pile-up	SN	1
Reconstructed Objects		
Electron trigger+reconstruction+ID+isolation	SN	4
Electron energy scale	SN	1
Muon trigger+reconstruction+ID+isolation	SN	12
Muon momentum scale+resolution	SN	3
Tau trigger+reconstruction+ID+EOR	SN	9
Tau scale	SN	3
Jet energy scale	SN	3
Jet energy resolution	SN	1
Jets JVT	SN	14
Jet Flavour Tagging	SN	23
Missing transverse momentum	SN	3
Background Modelling		
Fake Tau Estimation	SN	7
$t\bar{t}$	SN	2
$Z\tau\tau$ + HF	SN	2

Table 9.1: The main sources of the systematic uncertainties. The number of components for each category are shown with whether they are included as shape and normalisation variations (SN) or if they are normalisation only (N).

Chapter 10

Results

In order to place constraints on the benchmark models, a statistical fitting procedure based on the RooStats [96] framework is used. For the resonant di-Higgs production both a scalar heavy Higgs in the narrow width approximation, as predicted by 2HDM models, and a spin-2 RS KK graviton are considered for masses between 260 GeV and 1 TeV, where for both models the heavy resonances decay into two 125 GeV SM Higgs bosons. The non-resonant pair production of Higgs bosons is also considered assuming SM couplings. The BDT output score for each signal model and mass point are used as the discriminating variables for the statistical analysis. The binning of these distributions is optimised such that the background statistical uncertainty in each bin is kept to below 20% (40%) of the signal fraction for the resonant SLT (LTT) signals and 40% for the non-resonant signals in both channels. For bins with no signal the statistical uncertainty is required to be below 1%. Additionally a minimum of 10 events are required to be in each bin. The $Z^0 \to \mu\mu+$ heavy flavour jets control region, described in Section 8.3, is also included in the fit as a single bin to derive a normalisation factor for the Z^0+bb, bc, cc background processes.

The compatibility of the observations in the signal regions with the SM background only hypothesis is assessed using a profile likelihood ratio test. The likelihood function is constructed as the product of Poisson probability terms,

$$\mathcal{L}(\mu, \boldsymbol{\theta}) = \prod_{i \in \text{bins}} \frac{\left(\mu s_i(\boldsymbol{\theta}) + b_i(\boldsymbol{\theta})\right)^{n_i}}{n_i!} e^{-\left(\mu s_i(\boldsymbol{\theta}) + b_i(\boldsymbol{\theta})\right)} \prod_j \frac{1}{\sqrt{2\pi}\sigma_{\theta_j}} e^{\frac{-\left(\tilde{\theta_j} - \theta_j\right)}{2\sigma^2_{\theta_j}}}, \tag{10.1}$$

where $s_i(\boldsymbol{\theta})$ and $b_i(\boldsymbol{\theta})$ represent the expected signal and background yields in each bin, and n_i is the number of data events in each bin. The signal strength, μ, is the parameter of interest and is defined as the ratio of the fitted signal cross-section times branching fraction to the signal cross-section times branching ratio predicted by the particular signal model assumption.

The dependence of all sources of systematics and statistical uncertainty on the signal and background modelling are described by a set of nuisance parameters (NPs), $\boldsymbol{\theta}$. These are parameterised by Gaussian priors, with mean $\tilde{\theta}_j$ and variance σ_{θ_j} determined from subsidiary experimental measurements, which constrain the NPs within uncertainties to their nominal values. The likelihood function, $\mathcal{L}(\mu, \boldsymbol{\theta})$, is therefore a function of the signal strength, μ, and the NPs, $\boldsymbol{\theta}$.

The nominal fit result is obtained by maximising the likelihood function with respect to all parameters. The profile likelihood test statistic is then constructed as

$$q_\mu = -2 \ln \mathcal{L}\left(\mu, \hat{\hat{\theta}}_\mu\right) / \mathcal{L}\left(\hat{\mu}, \hat{\theta}\right), \tag{10.2}$$

where $\hat{\mu}$ and $\hat{\theta}$ are the parameters that maximise the likelihood, and $\hat{\hat{\theta}}_\mu$ are the values of the nuisance parameters that maximise the likelihood for a given μ. This statistic is used to test the compatibility of the observed data with the SM background only hypothesis, extracting the local p_0 value. For upper limits the test statistic takes the values

$$q_\mu = \begin{cases} -2 \ln \mathcal{L}\left(\mu, \hat{\hat{\theta}}_\mu\right) / \mathcal{L}\left(\hat{\mu}, \hat{\theta}\right) & \hat{\mu} \leq \mu, \\ 0 & \hat{\mu} > \mu, \end{cases} \tag{10.3}$$

as greater incompatibility between the hypothesised value of μ and the data result in higher values of q_μ, and data with $\hat{\mu} > \mu$ are not regarded as representing less compatibility so are not included as part of the rejection region of the test.

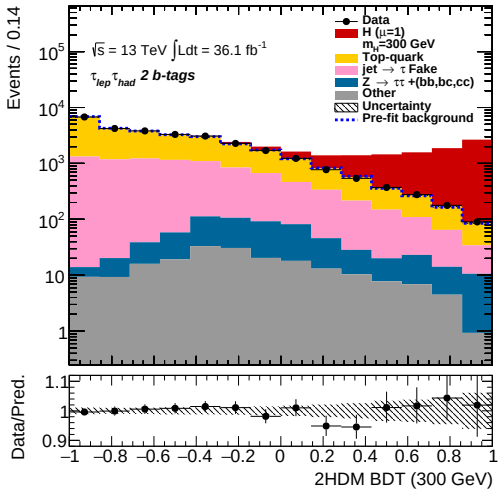
During the development of the analysis, specifically the selection and MVA optimisation, the data in the high score region of the final MVA distributions were blinded to avoid biasing the result. The MC modelling of the data for the kinematic variables and the inputs to the MVAs was validated in the 2 b-tag region after preselection. The fake tau background estimation was validated for the MVA output distributions in the 1 b-tag and same sign CRs. Once the background modelling was sufficiently well understood with reasonable agreement between data and MC the analysis was unblinded.

The observed data are found to be in good agreement with the predicted background events with no significant excesses. As such the data are used to set upper limits on the cross-section times branching ratio for the resonant di-Higgs production models and on the cross-section compared to the SM prediction for the non-resonant Higgs pair-production at the 95% confidence level (CL). The results are obtained from the profile likelihood ratio test using the CL_s prescription [97], where both the exclusion limits and p_0 values are

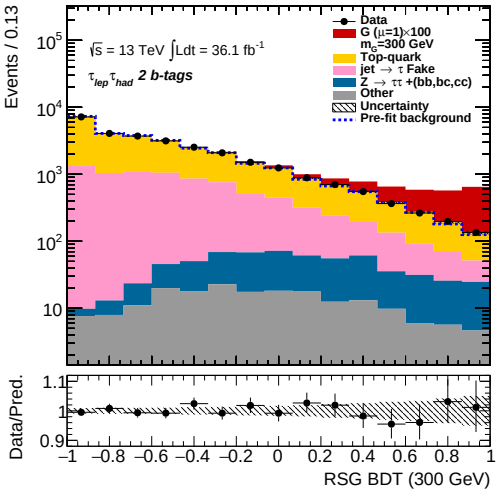
calculated using the asymptotic approximation [98].

The expected number of signal and background events after the fit has been performed for the region where the preselection criteria have been applied and exactly two b-tagged jets are selected are given in Table 10.1. The post-fit BDT distributions for the 2HDM and graviton searches when assuming a background only hypothesis are shown in Figures 10.1 and 10.2 for the SLT and LTT channels respectively. The non-resonant post-fit BDT distributions for both channels are shown in Figure 10.3.

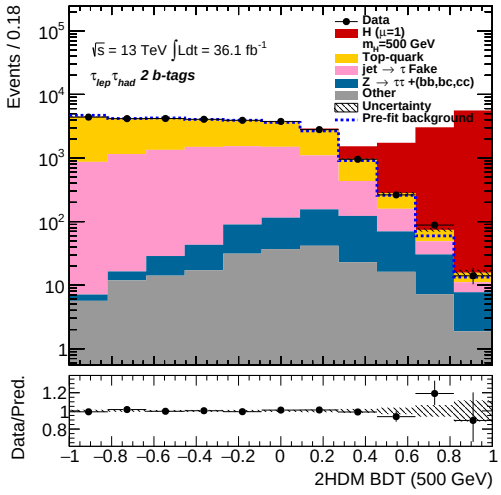
The results of the analysis for the resonance searches are presented as upper exclusion limits on each models' Higgs pair production cross-section times branching ratio to the $bb\tau\tau$ final state as a function of the resonance mass. The expected limits for the SLT, LTT and their combination for the 2HDM and graviton signals and the observed limit from the combination are shown in Figures 10.4 and 10.5 respectively. Upper limits on the non-resonant di-Higgs production cross-section are presented as multiples of the SM prediction in Table 10.2. The fractional impact of the the nuisance parameter on the error on the best fit signal strength $\hat{\mu}$ are shown in Table 10.3, calculated by removing each nuisance parameter in turn and repeating the fit comparing the now reduced post-fit uncertainty to the post-fit uncertainty obtained with all nuisance parameters included. These show that the uncertainty on the SM analysis is currently statistically dominated with the largest systematic uncertainties coming from the $t\bar{t}$ modelling, MC statistics and tau related detector uncertainties.



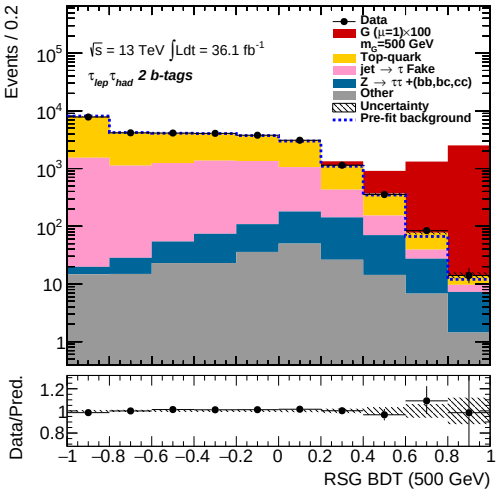
(a) 2HDM ($m_H = 300\,\text{GeV}$)



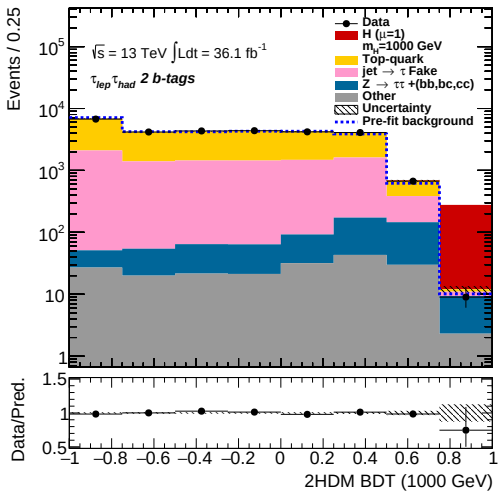
(b) RSG ($m_G = 300\,\text{GeV}$)



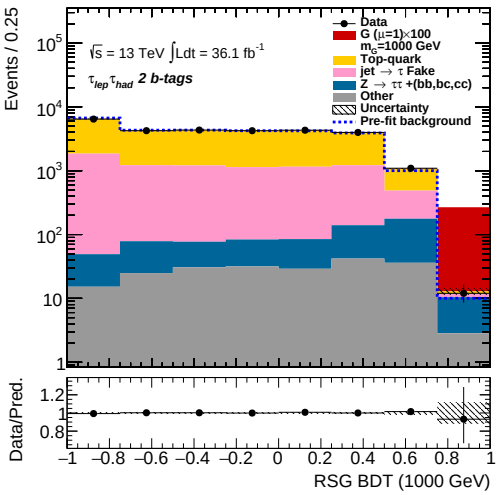
(c) 2HDM ($m_H = 500\,\text{GeV}$)



(d) RSG ($m_G = 500\,\text{GeV}$)

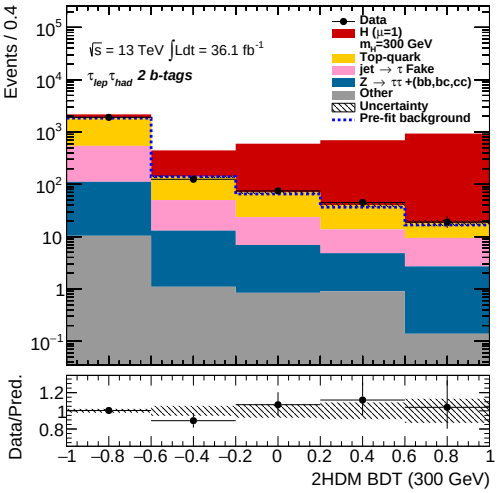


(e) 2HDM ($m_H = 1000\,\text{GeV}$)

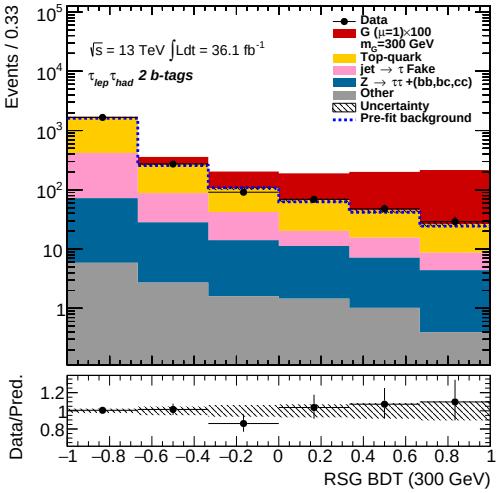


(f) RSG ($m_G = 1000\,\text{GeV}$)

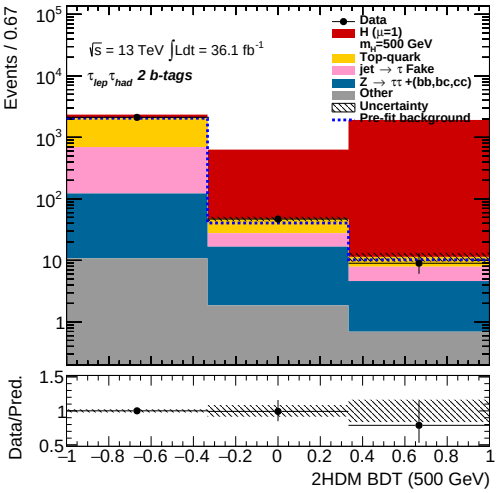
Figure 10.1: BDT output distributions using the optimised binning and after performing the fit, assuming a background only hypothesis, for the 2HDM (left) and RSG (right) signals in the $\tau_\ell\tau_{\text{had}}$ SLT channel for resonance masses of 300 GeV (top), 500 GeV (middle) and 1000 GeV (bottom).



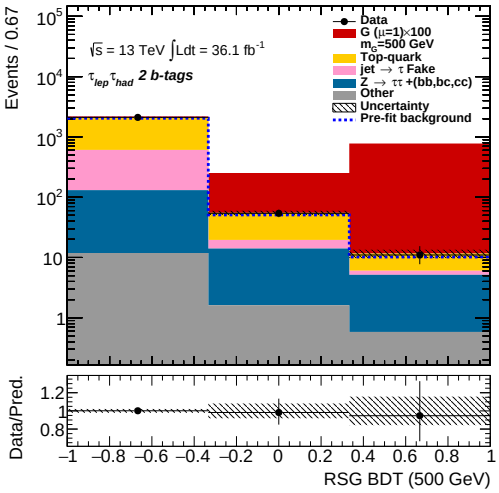
(a) 2HDM ($m_H = 300\,\text{GeV}$)



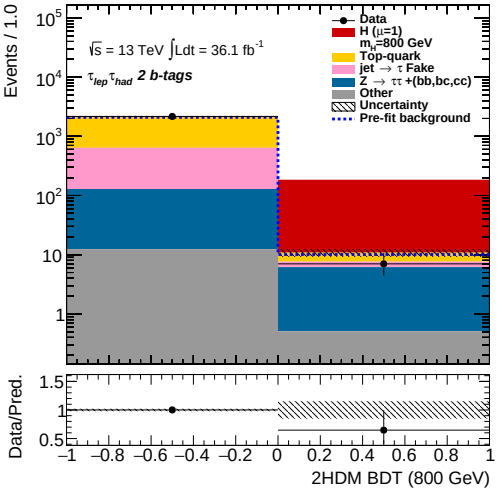
(b) RSG ($m_G = 300\,\text{GeV}$)



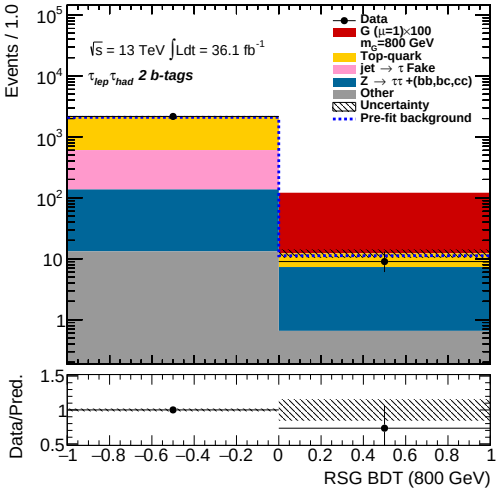
(c) 2HDM ($m_H = 500\,\text{GeV}$)



(d) RSG ($m_G = 500\,\text{GeV}$)



(e) 2HDM ($m_H = 800\,\text{GeV}$)



(f) RSG ($m_G = 800\,\text{GeV}$)

Figure 10.2: BDT output distributions using the optimised binning and after performing the fit, assuming a background only hypothesis, for the 2HDM (left) and RSG (right) signals in the $\tau_\ell\tau_{\text{had}}$ LTT channel for resonance masses of 300 GeV (top), 500 GeV (middle) and 800 GeV (bottom).

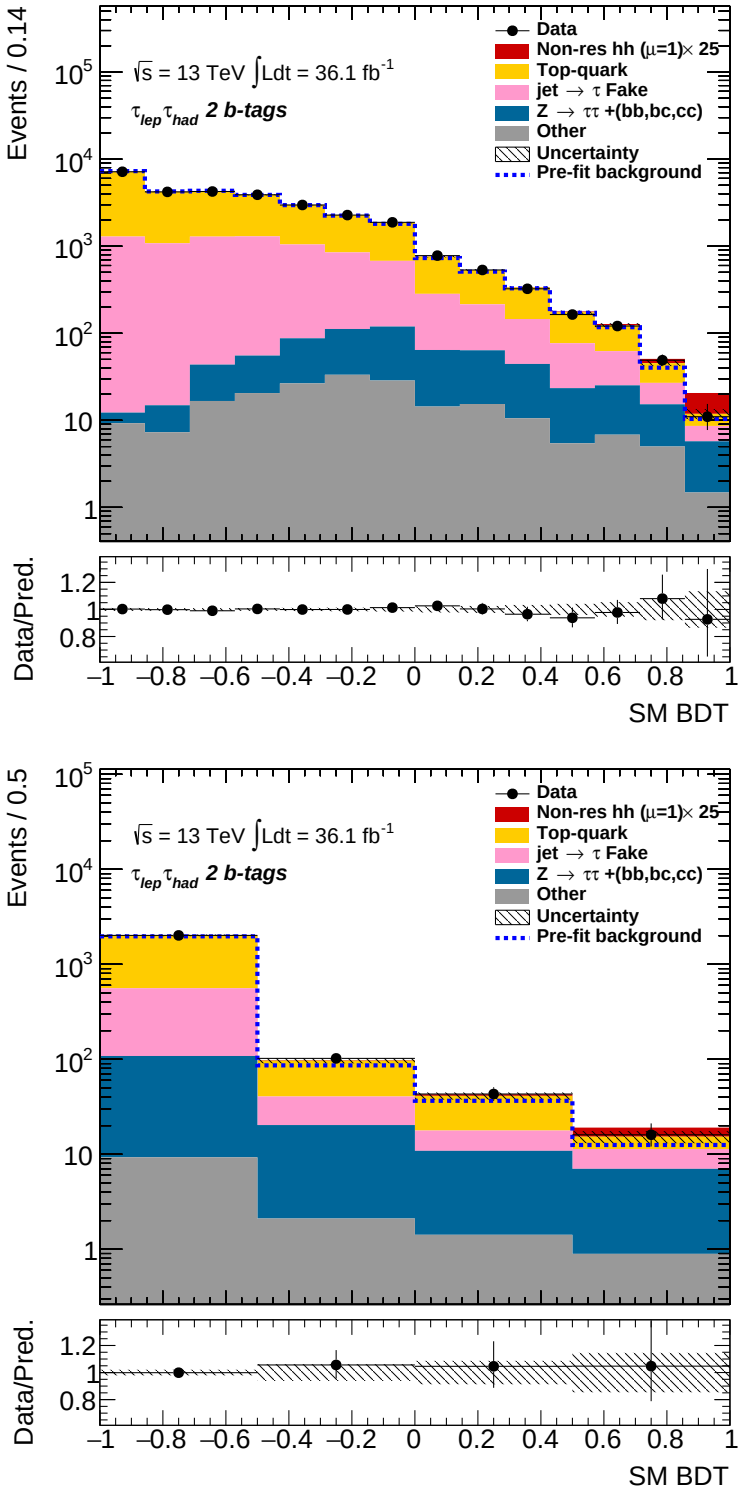


Figure 10.3: BDT output distributions using the optimised binning and after performing the fit, assuming a background only hypothesis, for the SM non-resonant Higgs pair production in the SLT (top) and LTT (bottom) channels.

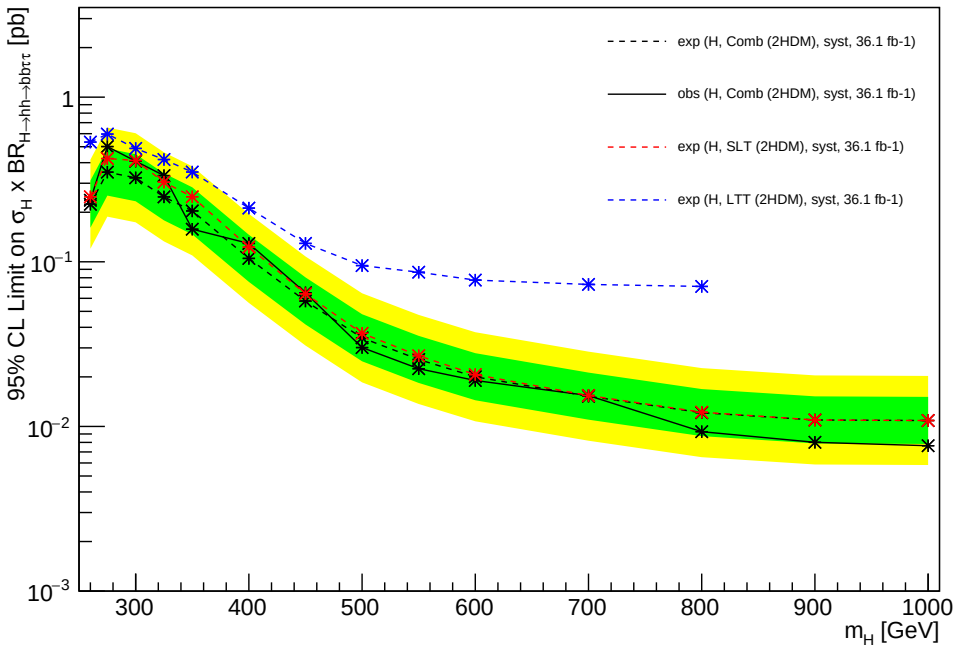


Figure 10.4: Expected (dashed black) and observed (solid black) limits at 95% C.L. on the cross-section times branching ratio of the 2HDM heavy scalar Higgs, $H \to hh \to bb\tau\tau$, process when combining the $\tau_\ell\tau_{\text{had}}$ SLT and LTT channels. The expected limits for the SLT (dashed red) and LTT (dashed blue) are also shown separately.

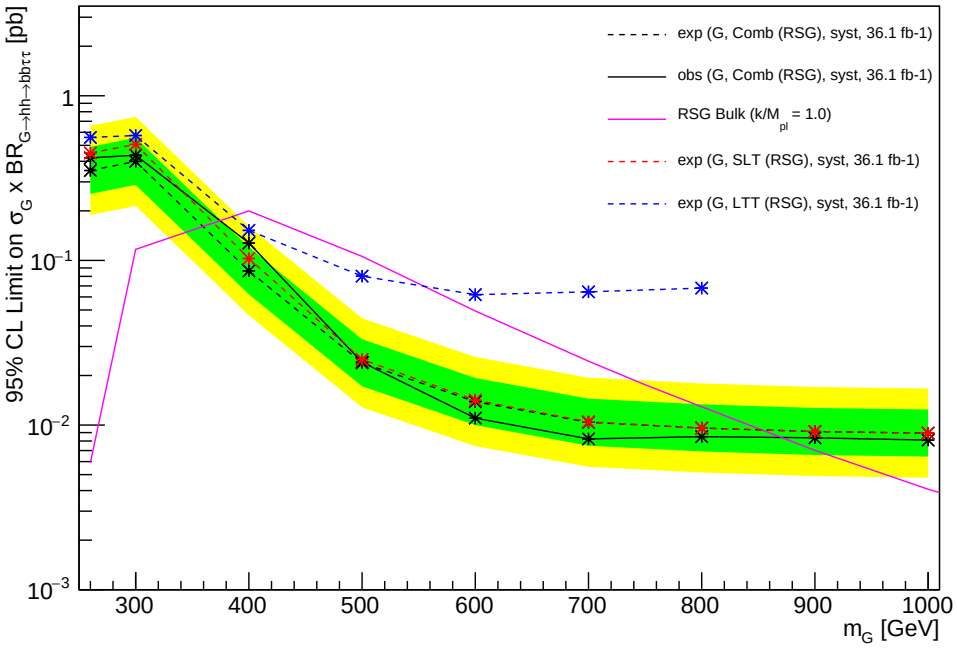


Figure 10.5: Expected (dashed black) and observed (solid black) limits at 95% C.L. on the cross-section times branching ratio of the KK RS graviton, $G \to hh \to bb\tau\tau$, process when combining the $\tau_\ell\tau_{\text{had}}$ SLT and LTT channels. The expected limits for the SLT (dashed red) and LTT (dashed blue) are shown separately as well as the expected benchmark cross-section for KK RS graviton production assuming k/M_{Pl} (magenta).

Sample	SLT	LTT
$G(m_G = 300\,\text{GeV},\, c = 1.0)$	22.8 ± 2.5	7.01 ± 0.74
$G(m_G = 500\,\text{GeV},\, c = 1.0)$	42.6 ± 4.1	9.85 ± 0.94
$G(m_G = 800\,\text{GeV},\, c = 1.0)$	7.8 ± 0.8	1.08 ± 0.10
$H(m_H = 300\,\text{GeV})$	9080 ± 1000 (38.4 ± 4.2)	2670 ± 300 (11.3 ± 1.3)
$H(m_H = 500\,\text{GeV})$	10500 ± 1000 (3.4 ± 0.3)	2670 ± 250 (0.86 ± 0.08)
$H(m_H = 800\,\text{GeV})$	1170 ± 110 (0.14 ± 0.01)	117 ± 17 (0.014 ± 0.002)
Non-res. $hh \to \tau_\ell \tau_h bb$	1.05 ± 0.10	0.29 ± 0.03
$t\bar{t}$	17600 ± 1200	1430 ± 150
Fake tau	9200 ± 1100	530 ± 140
Single top	1090 ± 110	68.2 ± 7.6
$Z^0 \to \tau\tau + (bb, bc, cc)$	464 ± 56	130 ± 15
Other $\left(W^\pm \,\&\, Z^0 \,\&\, VV\right)$	213 ± 16	18.3 ± 3.3
SM $Vh \to \tau\tau bb \,\&\, bb\tau\tau$	4.6 ± 2.4	1.04 ± 0.54
Total SM Background	28620 ± 190	2169 ± 46
Data	28613	2171

Table 10.1: The expected number events in the signal region where the preselection and 2 b-tagged jets are required. The first half of the table shows the expected number of signal events after the fit is performed with the signal strength fixed to $\mu = 1$. The yields for the 2HDM heavy scalars setting $\tan\beta = 2$ are shown in brackets as the yields for the benchmark model are found to be unrealistic and correspond to a previously excluded region of phase-space. The post-fit expected yields for the backgrounds assuming a background only hypothesis are shown in the second half of the table. The $t\bar{t}$ background includes events with true hadronic taus and a small contribution from events where a lepton fakes a hadronic tau. The 'Fake tau' background refers to any process where a jet fakes a hadronically decaying tau. The 'Other' category includes contributions from the $W^\pm$+jets, Z^0+ light flavour jets and diboson processes. The observed number of data events is also shown.

		Observed	-2σ	-1σ	Expected	$+1\sigma$	$+2\sigma$
SLT	σ [pb]	0.83	0.46	0.62	0.87	1.21	1.62
	$\sigma/\sigma_{\text{SM}}$	24.74	13.91	18.68	25.92	36.08	48.36
LTT	σ [pb]	2.86	1.37	1.84	2.55	3.55	4.77
	$\sigma/\sigma_{\text{SM}}$	85.55	41.03	55.09	76.45	106.40	142.63
Combined	σ [pb]	0.79	0.44	0.59	0.82	1.14	1.52
	$\sigma/\sigma_{\text{SM}}$	23.77	13.12	17.61	24.44	34.02	45.60

Table 10.2: Upper limits at the 95% CL on the non-resonant di-Higgs production cross-section for the SLT and LTT channels and their combination. The values are also compared to the predicted SM di-Higgs cross-section.

Set of nuisance parameters	Fractional impact on $\Delta\hat{\mu}$	
	Up/Down	Symmeterised
Data Statistics	$^{+0.66}_{-0.70}$	± 0.68
Systematic Uncertainties	$^{+0.34}_{-0.30}$	± 0.32
Normalisation	$^{+0.170}_{-0.093}$	± 0.13
Jets & $E_{\mathrm{T}}^{\mathrm{miss}}$	$^{+0.020}_{-0.012}$	± 0.016
b-tagging	$^{+0.062}_{-0.027}$	± 0.044
Electrons & Muons	$^{+0.00054}_{-0.00083}$	± 0.00067
Tau	$^{+0.094}_{-0.052}$	± 0.073
Luminosity	$^{+0.0043}_{-0.0010}$	± 0.0025
Fake estimation	$^{+0.0087}_{-0.011}$	± 0.0099
Top Modelling	$^{+0.180}_{-0.096}$	± 0.14
$Z\tau\tau$ + HF Modelling	$^{+0.011}_{-0.020}$	± 0.015
MC Statistics	$^{+0.078}_{-0.130}$	± 0.099

Table 10.3: The fractional impact on the error of the best fit signal strength, $\Delta\hat{\mu}$, of the nuisance parameters for the non-resonant signal. The impact is calculated by removing each nuisance parameter in turn and repeating the fit, comparing the now reduced post-fit uncertainty to the total post-fit uncertainty.

The results for the RS graviton interpretation exclude at the 95% CL gravitons in the mass range $380 < m_G < 870$ where it is assumed $c = k/M_{\mathrm{Pl}} = 1.0$. The limits have also been interpreted with $c = 2.0$, increasing the curvature of the warped space-time in the model. Under this assumption gravitons were excluded across the entire mass range, $260 < m_G < 1000$. This implies that future searches for RS gravitons considering this mass range should probe models with smaller values of c, increasing the radius of the extra curved dimension as c decreases. Searches for resonances with larger masses could also be conducted as these are currently not ruled out by this result.

The 2HDM signals were centrally generated by ATLAS with an agreed set of benchmark points. These are used to enable direct comparisons with previous results, however these benchmark points are from what is known to be an unrealistic region of the parameter space. The results corresponding to these require additional interpretation where specific model and parameters of those models need to be chosen. One such model is the minimal supersymmetric model scenario hMSSM [99] where exclusion regions can be set in the $(\tan\beta, m_A)$ plane, as was done for the Run 1 analysis shown in Figure 10.6. This interpretation is currently on-going with the updated exclusion limits from this analysis. The expected yields for the 2HDM signal points are shown renormalised with the cross-section calculated for $\tan\beta = 2$ in Tables 7.3 and 10.2, as this was not excluded by the Run 1 result and give a more reasonable number of expected events.

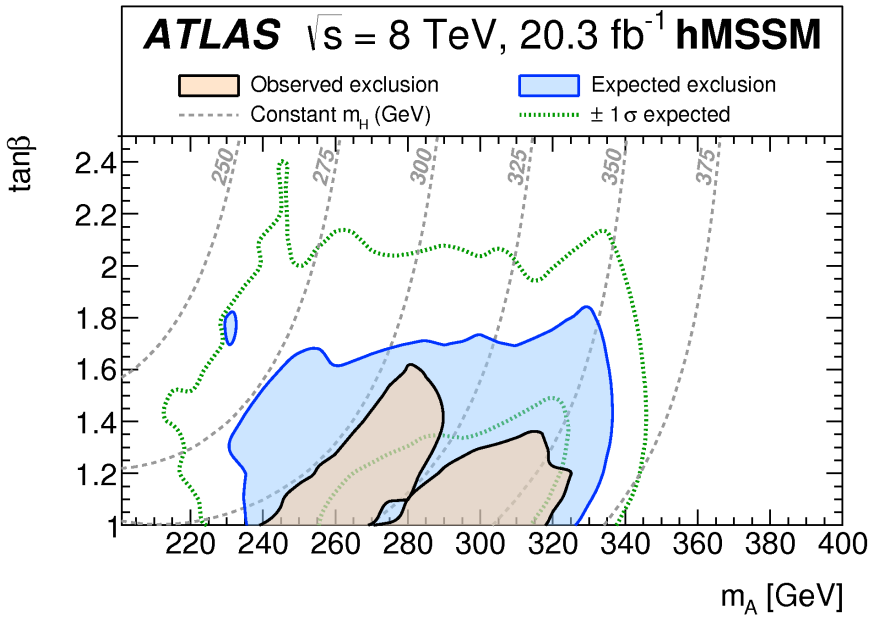


Figure 10.6: Expected and observed 95% CL exclusion regions in the $(\tan\beta, m_A)$ plane for the hMSSM scenario, from the Run 1 di-Higgs analysis [11].

Chapter 11

Multi-Variate Analysis

The nominal analysis makes use of BDTs to provide better separation between the signals and background than a purely cut based analysis would. Their use has recently been favoured in the HEP community due to their transparency and relative simplicity to optimise and train compared to some other MVA methods. However in some circumstances, such as lack of training MC, the BDT can be susceptible to overfitting. The nominal training for this analysis is also only performed on the $t\bar{t}$ background which leads to the $Z^0 \to \tau^+\tau^-+$ heavy flavour jets background being poorly classified as it is topologically more similar to the signal than the top background. This chapter explores the use of the SVM algorithm as well as separate trainings for the $t\bar{t}$ and $Z^0 \to \tau^+\tau^-+$ backgrounds to improve the results of the analysis.

11.1 Support Vector Machines

In the wider ML community SVMs are generally used where the training statistics are limited to give a more robust training compared to other methods, exhibiting less over-training [34][35, p.325–344][36, p.417–455]. This is particularly suited to the LTT channel of the analysis where the MC sample sizes for training are very small. This channel has only ~ 6000 MC events for the $t\bar{t}$ background and between approximately 5000 and 8000 MC events depending on the model and mass point compared to the SLT with ~ 80000 events for $t\bar{t}$ and 10s of thousand for the signals.

For both the resonant and non-resonant searches the SVMs are optimised and trained following the CV procedure described in Section 3.3.2. The samples are initially split into two folds based on odd and even event number. The first of these is randomly split in half again, one half is used for variable selection and the other for parameter optimisation, that are recombined to train the final SVM. The second fold is retained for evaluating the trainings later. The various kernel functions were compared when training the SVM and the

multi-gaussian, Equation 3.55, was found to give the best performance so is used throughout.

The variables were selected by training different combinations of pairs, with the highest performing pair, in terms of the area under the ROC curve (AUC), taken. Additional variables where then selected following the same process where all combinations were trained and the variable giving the largest improvement was added. The procedure was terminated when the gain in the AUC was $< 0.5\%$. The variables chosen for the resonant and non-resonant SVM trainings are shown in Table 11.1 with the variables used by the BDT.

Once the variables were selected, the hyper-parameters, C and Γ_i, of the SVM were tuned using the built-in parameter optimisation in TMVA, newly developed for the SVM. This uses the Minuit toolkit [100] to select the parameters which maximise the separation, defined as

$$\left\langle S^2 \right\rangle = \frac{1}{2} \int \frac{\left(\hat{y}_s(y) - \hat{y}_b(y)\right)^2}{\hat{y}_s(y) + \hat{y}_b(y)} \mathrm{d}y. \tag{11.1}$$

where $\hat{y}_s$ and $\hat{y}_b$ are the probability density functions (PDFs) of the signal and background responses for classifier y.[1]
The optimised hyper-parameters for the resonant and non-resonant SVMs are summarised in Appendix C.

Variables	BDT	Non-resonant SVM	Resonant SVM
m_{hh}	✓	✓	✓
$m_{\tau_\ell \tau_h}^{\mathrm{MMC}}$	✓		✓
$m_{\tau_\ell \tau_h}^{\mathrm{vis}}$		✓	
m_{bb}	✓	✓	✓
$\Delta R\,(l, \tau)$	✓	✓	✓
$\Delta R\,(b_1, b_2)$	✓	✓	✓
m_{T}^{W}	✓	✓	
p_{T}^{bb}			✓

Table 11.1: Variables used to train the BDTs and SVMs for the resonant and non-resonant analyses for the LTT channel. $m_{\tau_\ell \tau_h}^{\mathrm{vis}}$ is the mass of the visible decay products of the di-tau system; the electron or muon and the hadronic tau. p_{T}^{bb} is the transverse momentum of the di-b-jet system. All other variables are described in Section 7.3.1.

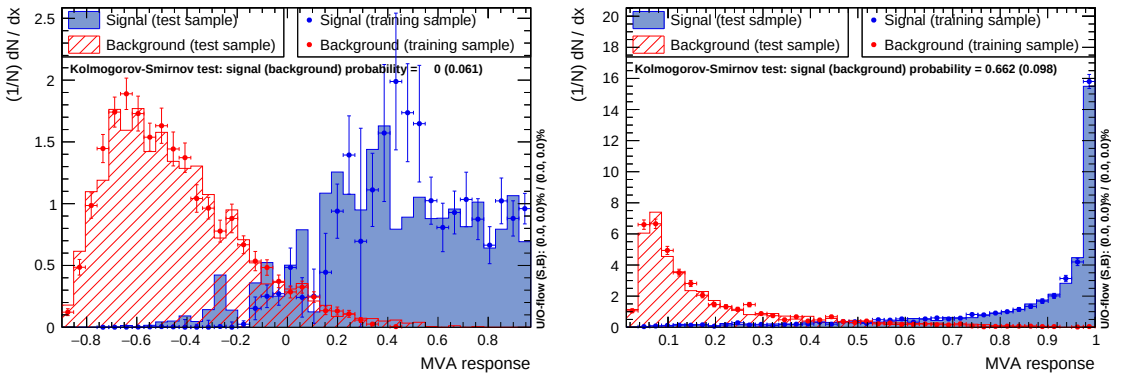
One notable difference between the SVMs and the BDTs is the time required to train

[1] The hyper-parameters of the MVA should ideally be optimised on the final FOM for the analysis, in this case the expected limit. However this is computationally challenging as would require propagating the resultant training for each set of parameters through the entire analysis chain and so a more simple metric is used to be more time efficient.

each method, with the BDT being relatively fast by comparison. For example to train the SVM on the LTT SM samples takes 11 seconds while the equivalent BDT takes less than 2 seconds. This is not too much of an issue for small samples such as these, however the training time for the SVM scales proportional to the number of data squared. This means that for the SLT analysis training a single SVM can take several hours, making the processes involving multiple trainings such as variable selection and parameter optimisation unfeasible. Another related issue is the memory usage of the computers required to calculate the $n \times n$ matrix of events for the SVM. One needs to use over 60 Gb of physical memory for the full SLT trainings making it only possible on some cluster computers.
Finally the application time of the SVMs for the analysis is related to the number of support vectors selected in the training increasing with the number of training events used as well as for increases in the cost parameter. This can be significantly longer than the application time of the BDTs, especially for analyses such as the resonant searches where multiple MVAs are evaluated, taking upwards of one hour per systematic variation for large samples such as the SLT channel. All of these factors demonstrate why the SVM is more suited to use on small data samples.

11.1.1 Non-resonant LTT SVM Results

The BDT and SVM were trained on the same set of samples for the non-resonant Higgs pair production signal and the $t\bar{t}$ background. Both methods obtain similar performance in terms of the AUC, however the BDT output distributions show some signs of overtraining for the signal samples with the training sample not describing the testing sample well in the low to mid BDT score region. The training for the SVM appears to be more robust with the training and testing samples in good agreement for both the signal and background.



(a) SM BDT (b) SM SVM

Figure 11.1: MVA output distributions for the SM di-Higgs signal trained against the $t\bar{t}$ in the LTT channel for the BDT (a) and the SVM (b). The BDT signal distribution exhibits some visible overtraining with some discrepancy between the training (points) and testing (histogram) samples. The SVM has good agreement between the training and testing samples for both the signal and background.

The trainings are passed through the full analysis chain with systematics to obtain upper exclusion limits on the non-resonant di-Higgs production cross-section, presented in Table 11.2.

		Observed	-2σ	-1σ	Expected	$+1\sigma$	$+2\sigma$
BDT	σ [pb]	2.86	1.37	1.84	2.55	3.55	4.77
	$\sigma/\sigma_{\mathrm{SM}}$	85.55	41.03	55.09	76.45	106.40	142.63
SVM	σ [pb]	4.11	2.19	2.94	4.07	5.67	7.60
	$\sigma/\sigma_{\mathrm{SM}}$	122.97	65.46	87.88	121.96	169.73	227.54

Table 11.2: Upper limits at the 95% CL on the non-resonant di-Higgs production cross-section in the LTT channel for comparing the nominal BDT and an optimised SVM. The values are also compared to the predicted SM di-Higgs cross-section.

Despite what appears to be a more robust training for the SVM a degradation of 59.5%(43.7%) in the expected(observed) limit is seen compared to the nominal BDT limit. This is due to a larger amount of the $Z^0 \to \tau\tau+$ heavy flavour jets background in the final bin of the SVM distrubution, shown in Figure 11.13, compared to the equivalent BDT, shown in Figure 10.3. This indicates that although the SVM achieves similar performance to the BDT when applied to the $t\bar{t}$ background, on which it is trained, that it performs less well on samples that it has not been trained on. A solution to this issue is explored later in this chapter.

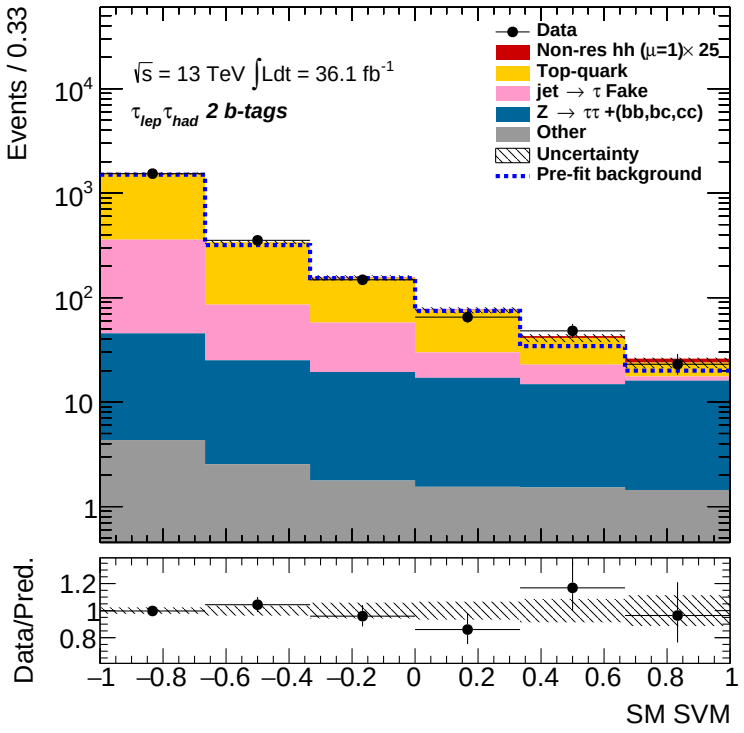


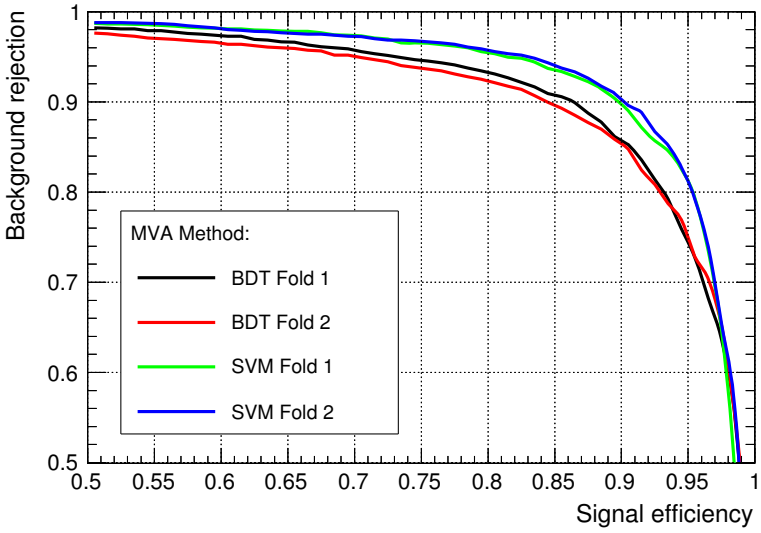
Figure 11.2: Post-fit SVM output distribution using the optimised binning, assuming a background only hypothesis, for the SM non-resonant Higgs pair production in the LTT channel.

11.1.2 Resonant LTT SVM Results

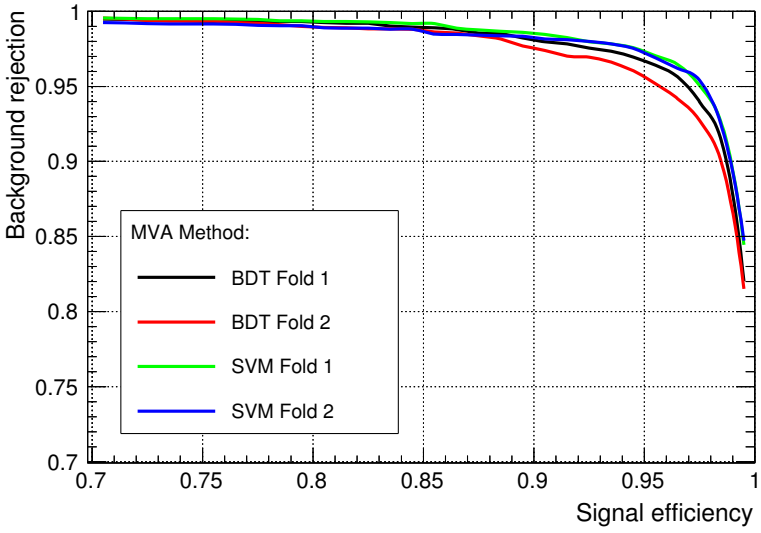
Similarly to the non-resonant, SVMs were also trained for the resonant 2HDM signal in the LTT channel following the procedure outlined above.
The ROC curves, shown in Figure 11.3, show slightly better performance by the SVM than the BDT in the low mass region, with both methods having similar discrimination for the rest of the mass range. The MVA output distributions are shown in Figure 11.4 and show reasonable agreement between the testing and training datasets in both cases.

The trainings were run through the full analysis chain with systematics to obtain upper exclusion limits for the 2HDM on the Higgs pair production cross-section times branching ratio to the $bb\tau\tau$ final state as a function of the resonance mass. The expected limits using SVMs were compared to those obtained using BDTs, presented in Figure 11.5, and show comparable performance between the two methods across the entire mass range, though the SVM is consistently slighty worse for the middle mass points. The observed limits compared to the expected limits for the results using the SVM are presented in Figure 11.6 and show no significant excesses. The post-fit SVM distributions, assuming a background only hypothesis are shown in Figure 11.7 with the corresponding BDT distributions for comparison.

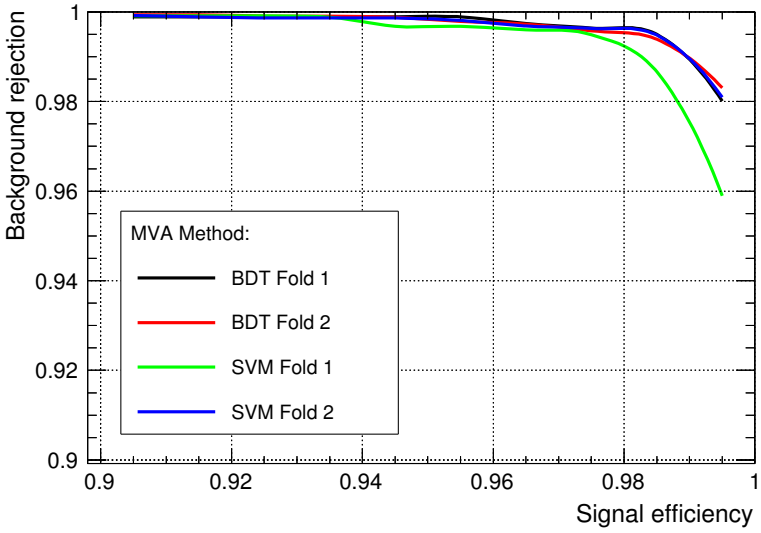
As no significant gains were found through the use of SVMs for the resonant analysis, the combination of the SLT BDT limits with LTT SVM limits was not performed as it would only yield similar results to the combination using only BDTs. These results however demonstrate that comparable results can be achieved to BDT based analyses using SVMs and so could benefit those analyses which are constrained by limited training statistics, potentially providing more robust trainings.



(a) 2HDM ($m_H = 300\,\text{GeV}$)

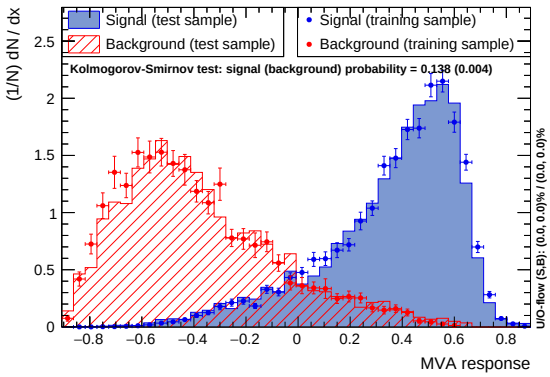


(b) 2HDM ($m_H = 500\,\text{GeV}$)

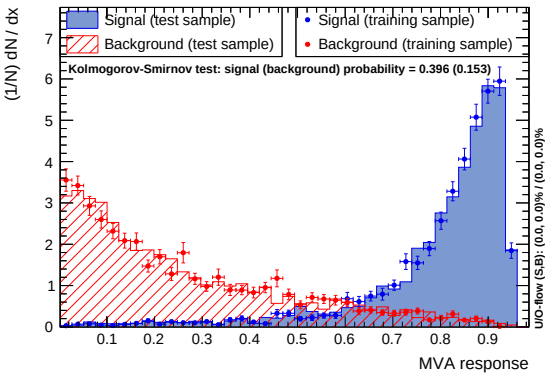


(c) 2HDM ($m_H = 800\,\text{GeV}$)

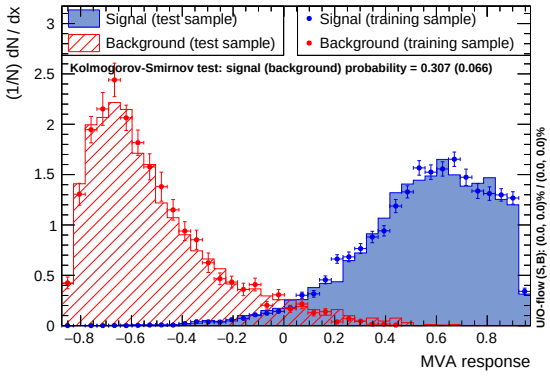
Figure 11.3: ROC curves comparing the BDTs to SVMs trained on the $t\bar{t}$ background and the 2HDM signal for the (a) 300 GeV, (b) 500 GeV and (c) 800 GeV mass points in the LTT channel.



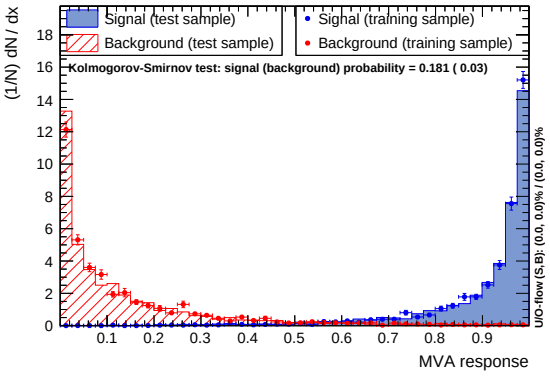
(a) 2HDM BDT ($m_H = 300\,\text{GeV}$)



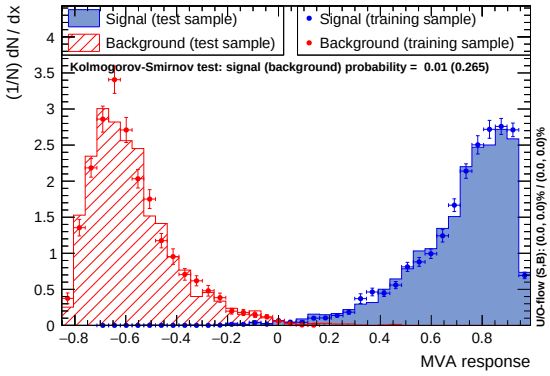
(b) 2HDM SVM ($m_H = 300\,\text{GeV}$)



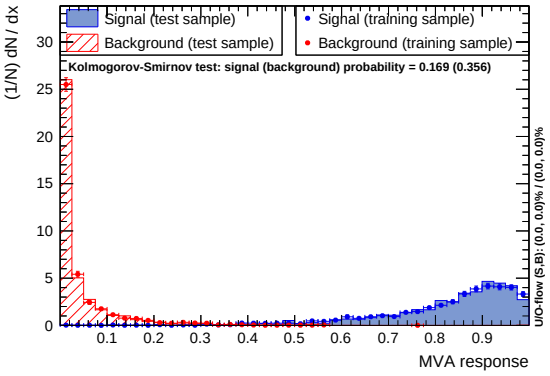
(c) 2HDM BDT ($m_H = 500\,\text{GeV}$)



(d) 2HDM SVM ($m_H = 500\,\text{GeV}$)



(e) 2HDM BDT ($m_H = 800\,\text{GeV}$)



(f) 2HDM SVM ($m_H = 800\,\text{GeV}$)

Figure 11.4: MVA output distributions comparing the BDTs (left) to SVMs (right) trained on the $t\bar{t}$ background and the 2HDM signal for the 300 GeV (top), 500 GeV (middle) and 800 GeV (bottom) mass points in the LTT channel.

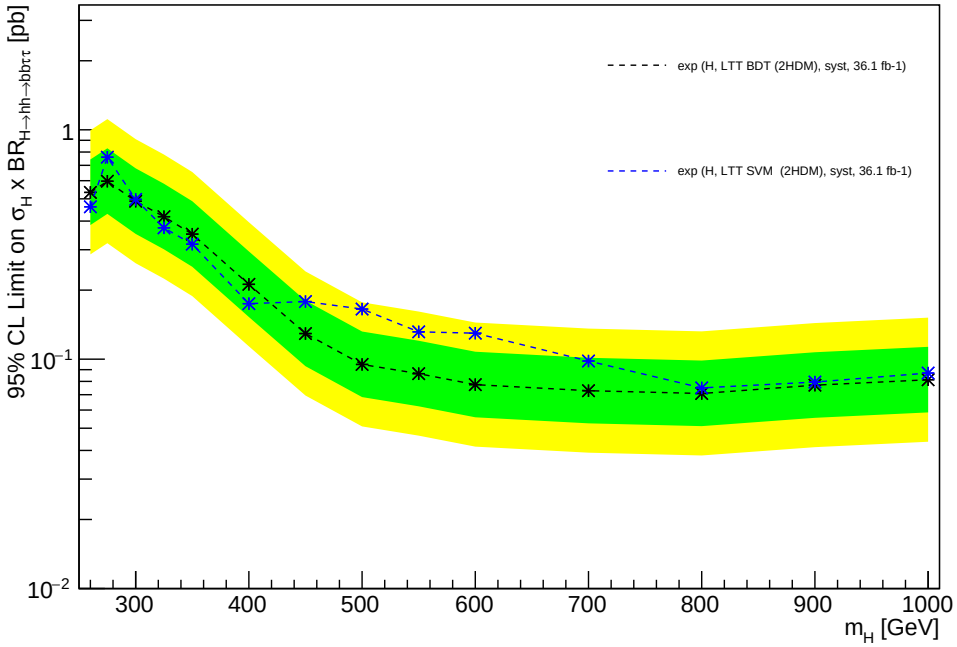


Figure 11.5: Expected limits for the BDT (black) and SVM (blue) at 95% C.L. on the cross-section times branching ratio of the 2HDM heavy scalar Higgs, $H \to hh \to bb\tau\tau$, process in the LTT channel.

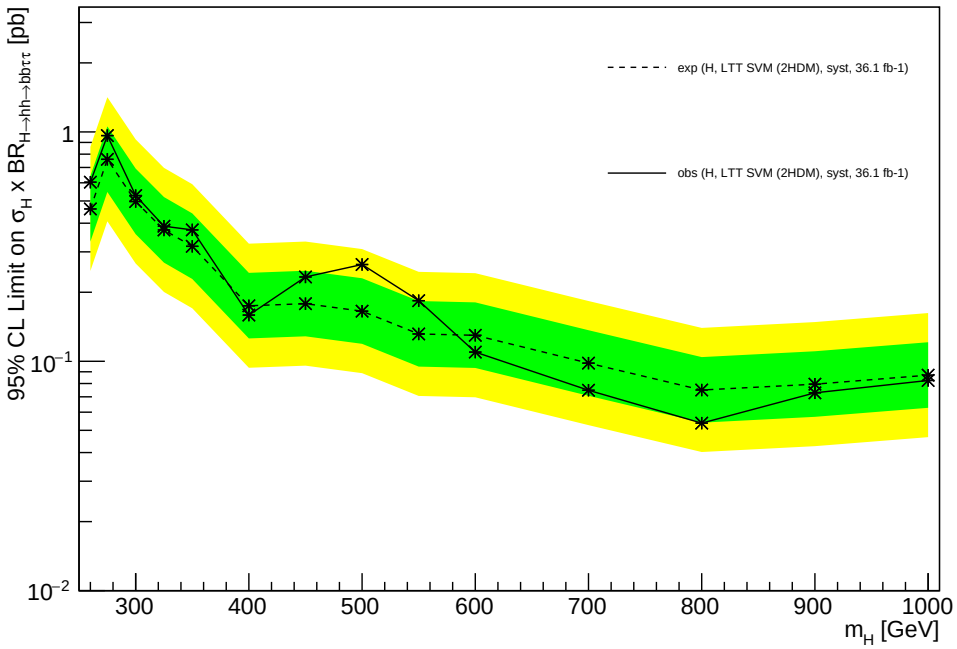
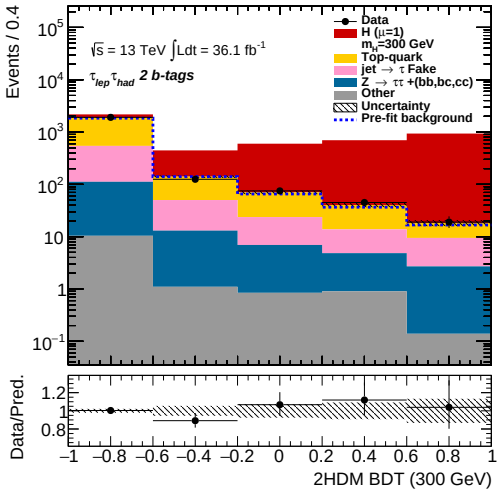
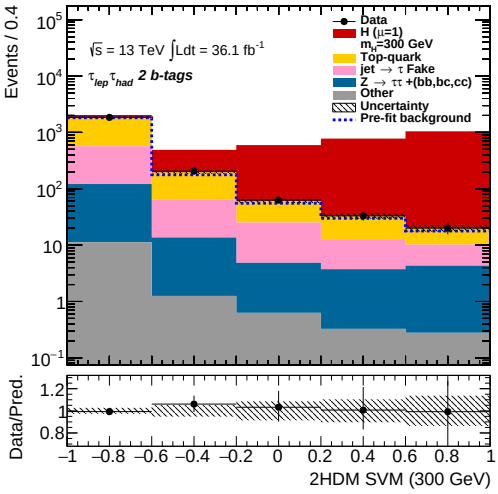


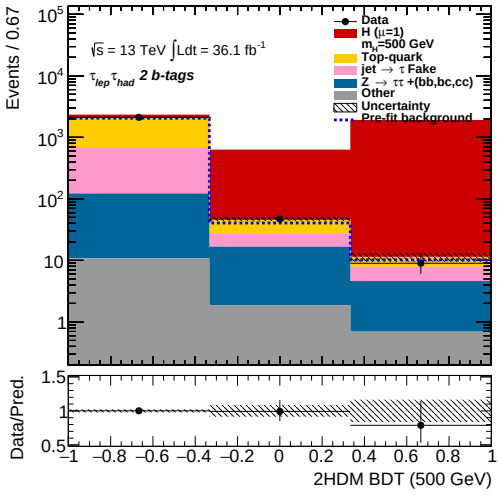
Figure 11.6: Expected (dashed black) and observed (solid black) limits using SVMs at 95% C.L. on the cross-section times branching ratio of the 2HDM heavy scalar Higgs, $H \to hh \to bb\tau\tau$, process in the LTT channel.



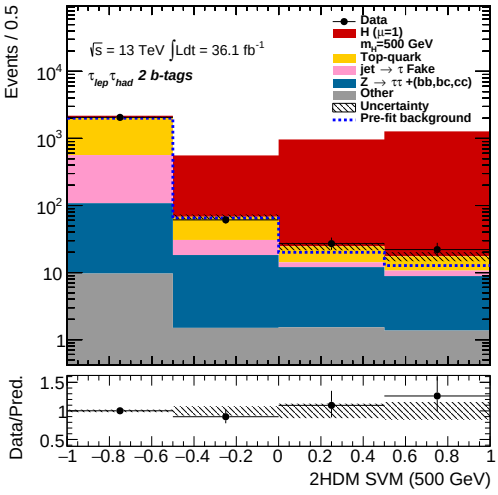
(a) 2HDM BDT ($m_H = 300\,\text{GeV}$)



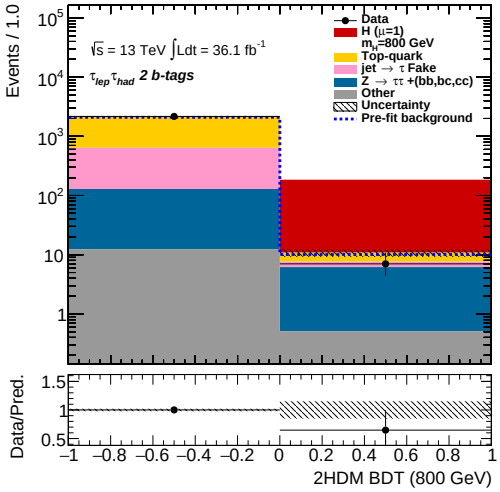
(b) 2HDM SVM ($m_H = 300\,\text{GeV}$)



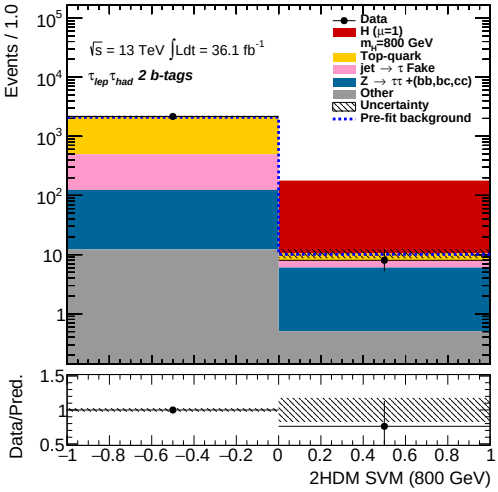
(c) 2HDM BDT ($m_H = 500\,\text{GeV}$)



(d) 2HDM SVM ($m_H = 500\,\text{GeV}$)



(e) 2HDM BDT ($m_H = 800\,\text{GeV}$)



(f) 2HDM SVM ($m_H = 800\,\text{GeV}$)

Figure 11.7: MVA output distributions using the optimised binning and after performing the fit, assuming a background only hypothesis, for the BDT (left) and SVM (right) in the $\tau_\ell\tau_{\text{had}}$ LTT channel for resonance masses of 300 GeV (top), 500 GeV (middle) and 800 GeV (bottom).

11.2 Committee Machine

For the non-resonant SLT analysis the events from $Z^0 \to \tau^+\tau^- +$ heavy flavour jets ($Z\tau\tau + \text{HF}$) accumulate in the high BDT region for the nominal training, specifically events containing two real b-jets. This is problematic as the MC does not model the data well for this background, being scaled up significantly in the final fit, and as such the associated systematics have a large impact. It was investigated as part of the nominal analysis if combining these background samples with the $t\bar{t}$ samples in the training of the BDT would lead to any additional separation and therefore an improvement in the limits, however this was not found to be the case as there seems to be some tension between the two background samples.

In order to evaluate the potential gain from improving the separation between the signal and the $Z\tau\tau + \text{HF}$ background, as a benchmark the final statistical-only limits were run excluding the $Z\tau\tau + \text{HF}$ samples from the fit, yielding an upper limit of $\sigma^{hh}_{\text{exp}}/\sigma^{hh}_{\text{SM}} = 18.80$ compared to the nominal $\sigma^{hh}_{\text{exp}}/\sigma^{hh}_{\text{SM}} = 20.57$. This shows a maximum improvement of 8.6% could be achieved if the $Z\tau\tau + \text{HF}$ background could be completely removed from the BDT distribution.

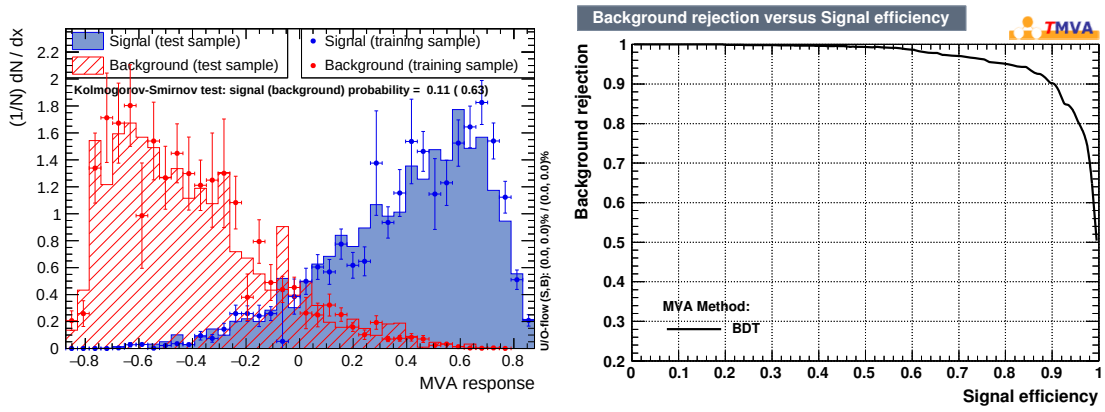
A potential way to achieve additional separation is to make use of committee machines [35, p.655–657] where expert learners are trained individually for the $t\bar{t}$ and $Z\tau\tau + \text{HF}$ backgrounds and then their outputs are combined in some way to produce the final discriminating variable. The nominal training for the SLT analysis is already performed with just the $t\bar{t}$ background, so an additional training is required for the $Z\tau\tau + \text{HF}$ background.

Variables	$t\bar{t}$ Training	$Z\tau\tau + \text{HF}$ Training
m_{hh}	✓	✓
$m^{\text{MMC}}_{\tau_\ell\tau_h}$	✓	✓
m_{bb}	✓	✓
$\Delta R\,(l, \tau)$	✓	✓
$\Delta R\,(b_1, b_2)$	✓	✓
m^W_{T}	✓	
$\Delta\eta\,(b, b)$		✓

Table 11.3: Optimised variables for the BDT trained separately on the $t\bar{t}$ and $Z\tau\tau + \text{HF}$ backgrounds against the non-resonant di-Higgs signal for the SLT channel.

The variables for the $Z\tau\tau + \text{HF}$ BDT training were selected following the same procedure described in Section 7.3.1 and are summarised in Table 11.3 with the nominal BDT variables for comparison. These are found to be similar to the nominal training, with m^W_{T} replaced by

$\Delta\eta\,(b,b)$ which provides better separation, and the same hyper-parameters are used for the BDT as for the nominal $t\bar{t}$ training. The BDT output distribution is shown in Figure 11.8(a) and the ROC curve, Figure 11.8(b), shows similar performance to the nominal BDT.



(a) (b)

Figure 11.8: (a) BDT output distribution and (b) ROC curve for non-resonant signal trained against the $Z\tau\tau$ + HF background in the SLT channel.

The output distributions of the two BDT trainings are shown compared to each other for the non-resonant di-Higgs signal and the $t\bar{t}$ and $Z\tau\tau$ + HF backgrounds in Figure 11.9. The simplest way to combine the two MVAs is via a linear average where the combined output for a given event is defined as

$$\tilde{y} = \alpha \times y_Z + (1-\alpha) \times y_{t\bar{t}} \tag{11.2}$$

where y_Z and $y_{t\bar{t}}$ are the outputs of the separate $Z\tau\tau$ + HF and $t\bar{t}$ trainings respectively and α is a parameter to be optimised that weights how much of each MVA is in the final classifier. α was selected by scanning through different values and evaluating the statistical-only expected limits to find the minimum, shown in Figure 11.10, and found to be $\alpha = 0.25$. Other methods of combining these outputs could be considered such as using them as inputs to another MVA, however the performance gained from the simple linear combination is already close to that of removing the $Z\tau\tau$+HF background altogether.

The committee MVA is run through the full analysis with systematics to set upper exclusion limits on the cross-section for non-resonant Higgs pair production using the nominal $t\bar{t}$ BDT, the $Z\tau\tau$ + HF BDT and the committee BDT, shown in Table 11.4. The committee method provides an improvement of 13.9% in the expected limit compared to the nominal SLT result. This is due to the reduction of the $Z\tau\tau$ + HF background in the high score region of the BDT distribution, shown in Figure 11.11, compared to the nominal BDT distribution, Figure 10.3. Due to an upward fluctuation in the data in the final bin the observed limit got worse by 16%.

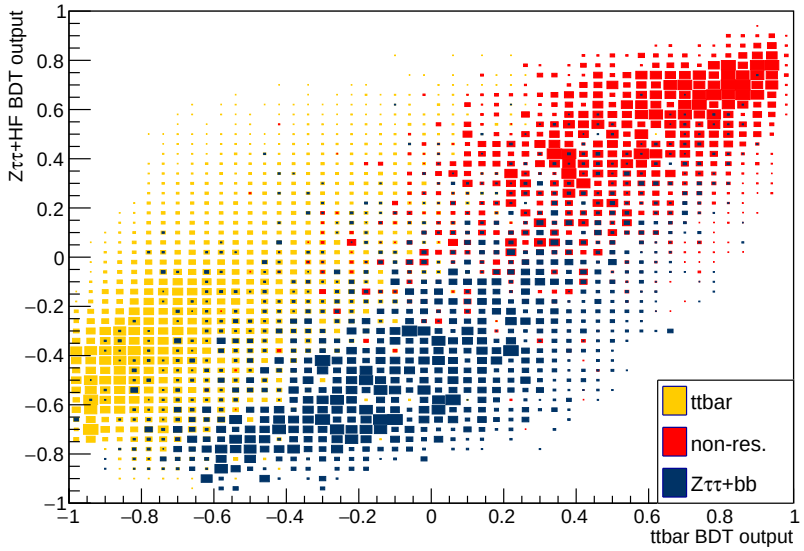


Figure 11.9: Output distribution for the BDT trained on the $t\bar{t}$ background against the output distribution for the BDT trained on the $Z\tau\tau$ + HF background for true tau $t\bar{t}$ (yellow) and $Z\tau\tau$ + HF (blue) background, and non-resonant SM signal (red) events.

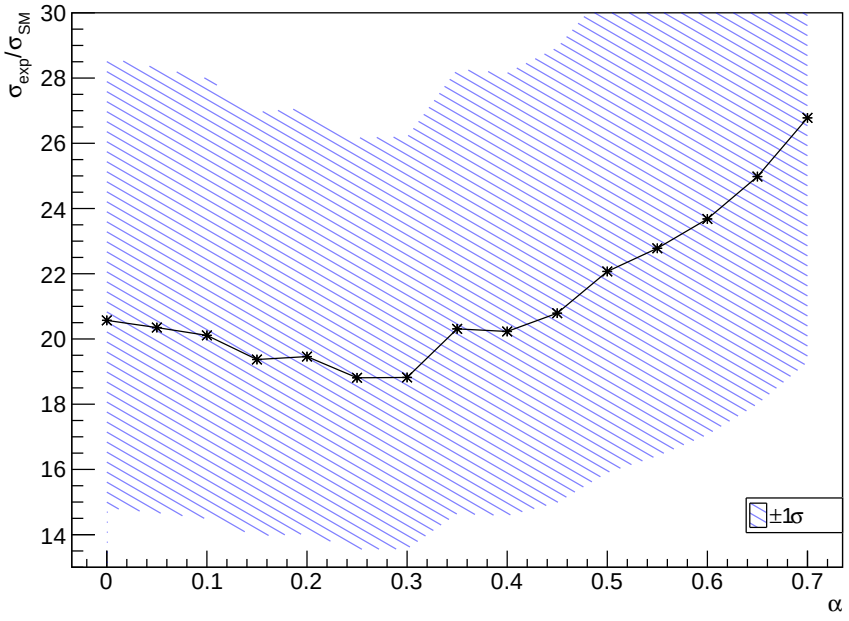


Figure 11.10: Optimisation scan of the α parameter, that weights the contribution of the $t\bar{t}$ and $Z\tau\tau$ + HF BDTs in the committee BDT, against the expect statistical only limit on the di-Higgs production cross-section compared to the SM prediction (black).

		Observed	-2σ	-1σ	Expected	$+1\sigma$	$+2\sigma$
$t\bar{t}$ BDT	σ [pb]	0.83	0.46	0.62	0.87	1.21	1.62
	$\sigma/\sigma_{\mathrm{SM}}$	24.74	13.91	18.68	25.92	36.08	48.36
$Z \to \tau\tau + \mathrm{HF}$ BDT	σ [pb]	1.98	1.00	1.34	1.87	2.60	3.48
	$\sigma/\sigma_{\mathrm{SM}}$	59.33	29.97	40.24	55.85	77.72	104.19
Committee BDT	σ [pb]	0.96	0.40	0.54	0.75	1.04	1.39
	$\sigma/\sigma_{\mathrm{SM}}$	28.71	11.99	16.09	22.33	31.08	41.67

Table 11.4: Upper limits at the 95% CL on the non-resonant di-Higgs production cross-section in the SLT channel for BDTs trained separately on the $t\bar{t}$ and $Z\tau\tau + \mathrm{HF}$ backgrounds only and the committee BDT.

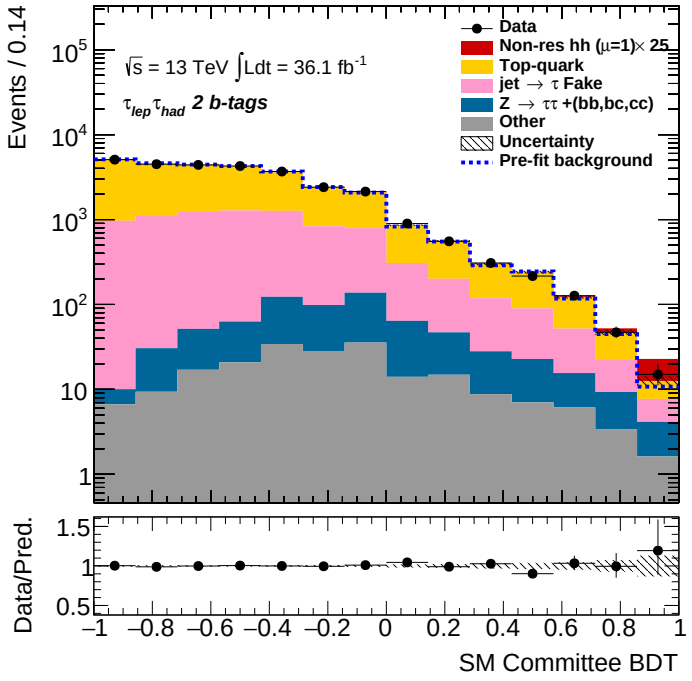


Figure 11.11: Post-fit committee BDT output distribution using the optimised binning, assuming a background only hypothesis, for the SM non-resonant Higgs pair production in the SLT channel.

11.3 SVM Committee Machine

The use of committee machines improved the expected result of the SLT non-resonant BDT analysis by reducing the $Z\tau\tau + \text{HF}$ background in the high score region of the BDT. The same methodology can be applied to improve the SVMs for LTT channel where it is particularly suited due to the low training statistics but suffers from the same problem, described in Section 11.1.1. The variables are re-optimised for the $Z\tau\tau + \text{HF}$ background following the same process as before and are shown in Table 11.5. The hyper-parameters are also optimised and are summarised in Appendix C.

Variables	$t\bar{t}$ SVM	$Z\tau\tau + \text{HF}$ SVM
m_{hh}	✓	✓
$m^{\text{MMC}}_{\tau_\ell\tau_h}$		✓
$m^{\text{vis}}_{\ell\tau}$	✓	
m_{bb}	✓	✓
$\Delta R\,(l,\tau)$	✓	✓
$\Delta R\,(b_1,b_2)$	✓	✓
p^{bb}_{T}		✓

Table 11.5: Optimised variables for the SVM trained separately on the $t\bar{t}$ and $Z\tau\tau + \text{HF}$ backgrounds against the non-resonant di-Higgs signal for the LTT channel.

The parameter determining the weighting of the two separate trainings in the committee is scanned between 0 and 1, Figure 11.12, and the statistical-only expected limit on the non-resonant di-Higgs production cross-section is used to determine the optimal point of $\alpha = 0.7$.

The full systematic analysis is run with the committee SVM. The post-fit SVM output distribution is shown in Figure 11.13. Upper limits are set on the non-resonant Higgs pair-production cross-section presented in Table 11.6. The result with the committee SVM improved by 42.1%(32.7%) for expected (observed) limit compared to the SVM trained on only the $t\bar{t}$ background, making it comparable to the limits achieved by the nominal BDT in the non-resonant LTT analysis slightly improving the expected (observed) limits by 7.6%(3.2%).

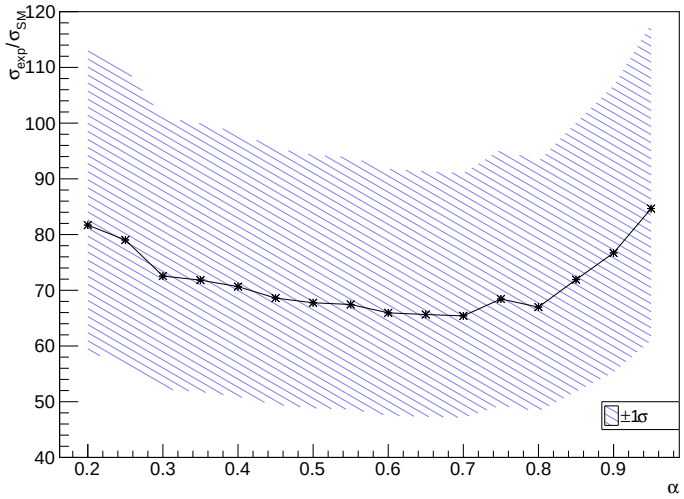


Figure 11.12: Optimisation scan of the α parameter, that weights the contribution of the $t\bar{t}$ and $Z\tau\tau + \text{HF}$ SVMs in the committee SVM, against the expect statistical only limit on the di-Higgs production cross-section compare to the SM prediction.

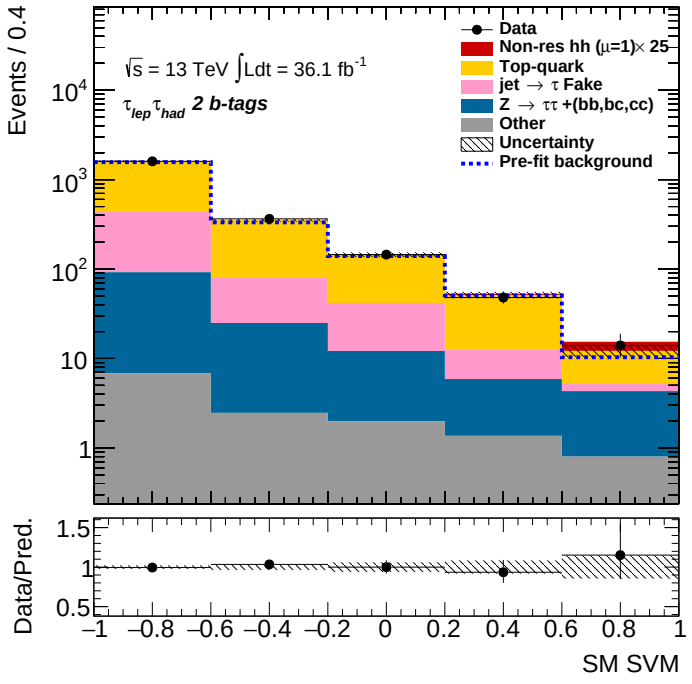


Figure 11.13: Post-fit committee SVM output distribution using the optimised binning, assuming a background only hypothesis, for the SM non-resonant Higgs pair production in the LTT channel.

		Observed	-2σ	-1σ	Expected	$+1\sigma$	$+2\sigma$
$t\bar{t}$ SVM	σ [pb]	4.11	2.19	2.94	4.07	5.67	7.6
	$\sigma/\sigma_{\mathrm{SM}}$	122.97	65.46	87.88	121.96	169.73	227.54
Committee SVM	σ [pb]	2.77	1.27	1.70	2.36	3.28	4.40
	$\sigma/\sigma_{\mathrm{SM}}$	82.80	37.91	50.90	70.63	98.30	131.78

Table 11.6: Upper limits at the 95% CL on the non-resonant di-Higgs production cross-section in the LTT channel for the SVM trained on the $t\bar{t}$ background only and the committee SVM.

Chapter 12

Summary and Outlook

A search for resonant and non-resonant Higgs pair production decaying to $bb\tau\tau$, where one tau decays hadronically and the other decays leptonically, is presented in this thesis. The analysis used 36.1fb^{-1} of data from p-p collisions recorded by the ATLAS experiment during 2015 and 2016 with centre-of-mass energy $\sqrt{s} = 13\,\text{TeV}$. Results from the non-resonant analysis are interpreted in the context of SM di-Higgs production and the resonant search is interpreted in the context of the production of heavy scalar Higgs in 2HDM models and spin-2 RS KK gravitons.

For the graviton interpretation of the analysis, upper exclusion limits were set on the hh production cross-section times branching ratio to $bb\tau\tau$ as a function of resonance mass at the 95% CL. Gravitons are excluded in the mass range $380 < m_G < 870\,\text{GeV}$ assuming $c = k/M_{\text{Pl}} = 1.0$. Future searches for RS gravitons should focus on models with smaller values of c or search for gravitons with higher masses. Upper limits are also set on the cross-section times branching as a function of m_H for the 2HDM interpretation.

Limits are set for the non-resonant di-Higgs production cross-section and interpreted compared to the SM prediction excluding cross-sections above $24.44\,\sigma^{hh}_{\text{SM}}$. This result is a significant improvement over the Run 1 ATLAS $bb\tau_{\text{lep}}\tau_{\text{had}}$ result of $\sim 160\,\sigma^{hh}_{\text{SM}}$, setting a limit more than 6 times better than that publication. The results are also stronger than the most recent limits set by CMS for this decay mode using 35.9fb^{-1} of data at $\sqrt{s} = 13\,\text{TeV}$ which includes $\tau_{\text{lep}}\tau_{\text{had}}$, $\tau_{\text{had}}\tau_{\text{had}}$ and boosted object categories.

The work contained within this thesis was conducted as part of a wider ATLAS analysis group working on the soon to be published analysis searching for di-Higgs to $bb\tau\tau$, also including the $bb\tau_{\text{had}}\tau_{\text{had}}$ channel which will bring further improvements. The author's contributions to this analysis included work towards the optimisation of the multi-variate analysis (particularly for the LTT channel), the development of the fake tau background estimation and its related systematics, and the combination of the SLT and LTT channels.

A boosted analysis which will improve and extend the high mass region of the resonant analysis is being explored as well as the VBF production of two Higgs bosons, which in certain areas of the 2HDM phase space can have an enhanced cross-section compared to the ggF production, both of which will potentially be included in the full Run 2 analysis.

The combination with other di-Higgs decay modes that have similar sensitivity, such as $bbbb$ and $bb\gamma\gamma$, is on-going. This could reach as low as $\sim 10\,\sigma^{hh}_{\mathrm{SM}}$ with the 2015 and 2016 dataset, gaining further improvements from the full Run 2 dataset. The combination will also set limits for the non-resonant di-Higgs production cross-section as a function of λ_{hhh} compared to the SM value, closing in on its future measurement. Work is also on-going to interpret the 2HDM result in the minimal supersymmetric model scenario hMSSM [99] setting exclusion limits in the $(\tan\beta, m_A)$ plane, similar to what was done in the Run 1 di-Higgs analysis.

The $bb\tau\tau$ analysis, similar to the other di-Higgs analyses, is currently statistically limited. However by the end of Run 2 and looking forward eventually to the HL-LHC, the impact of the systematics will become an increasing problem to achieving improvements through the addition of new data to the analysis. New ideas and methods are required to overcome this and bring the analyses closer to the observation of di-Higgs production.

In an attempt to achieve this goal several candidates to enhance the separation of signal and background from the MVAs were explored in this thesis. One promising method is training committees of MVAs, each expert in suppressing a particular background, whose combined output gives better discrimination than training on just one or all of the background together, as demonstrated for the non-resonant SLT analysis.

Support vector machines were applied to the SM non-resonant and 2HDM resonant LTT channel searches where the limited statistics made them an ideal algorithm to try. Training following a similar methodology to the nominal BDT yielded comparable results for the resonant signal (in terms of the limits on the di-Higgs cross-section), while for the non-resonant the results were worse compared to the BDT. The SVM however showed less signs of visible overtraining than the BDT with these limited training statistics, making it a good MVA candidate for analyses which lack sufficiently sized training samples to utilise other popular MVA methods. For the SM non-resonant analysis the combination of SVMs trained for the $t\bar{t}$ and $Z\tau\tau + \mathrm{HF}$ backgrounds in a committee slightly improved upon the nominal BDT result.

As the LHC approaches the end of Run 2 the prospect of physics beyond the standard model still eludes us, however the unprecedented centre-of-mass energy and luminosity allows for the refinement of the searches probing wider ranges of phase space. The study of extremely rare SM processes is also becoming possible with the quantity of data now being produced, helping to further constrain and confirm the SM itself. The search for di-Higgs production will be a key contributor to this, constraining the Higgs boson self-coupling and shape of the Higgs potential. It is not thought to be possible to observe this process until the end of the HL-LHC, but potential improvements such as those presented here bring us ever closer.

In conclusion, why make life easy looking for one Higgs at a time when you can look for two!

Bibliography

[1] G. Aad et al. [The ATLAS Collaboration]. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Physics Letters B*, 716(1):1–29, 2012.

[2] S. Chatrchyan et al. [The CMS Collaboration]. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Physics Letters B*, 716(1):30–61, 2012.

[3] G. Aad et al. [The ATLAS Collaboration]. Study of the spin and parity of the Higgs boson in diboson decays with the ATLAS detector. *The European Physical Journal C*, 75(10):476, 2015.

[4] S. Chatrchyan et al. [The CMS Collaboration]. Measurement of the properties of a Higgs boson in the four-lepton final state. *Physical Review D*, 89:092007, 2014.

[5] The CMS Collaboration] G. Aad et al. [The ATLAS Collaboration. Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments. *Physical Review Letters*, 114(19):191803, 2015.

[6] D. de Florian et al. [LHC Higgs Cross Section Working Group]. Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector. *CERN Yellow Reports: Monographs*, 2, 2017.

[7] L. Fitzpatrick, J. Kaplan, L. Randall, and L.-T. Wang. Searching for the Kaluza-Klein graviton in bulk RS models. *Journal of High Energy Physics*, 2007(09):013, 2007.

[8] K. Agashe, H. Davoudiasl, G. Perez, and A. Soni. Warped gravitons at the CERN LHC and beyond. *Phys. Rev. D*, 76:036006, 2007.

[9] G.C. Branco, P.M. Ferreira, L. Lavoura, M.N. Rebelo, M. Sher, and J. P. Silva. Theory and phenomenology of two-Higgs-doublet models. *Physics Reports*, 516(1):1–102, 2012.

[10] J. F. Gunion, S. Dawson, H. E. Haber, and G. L. Kane. *The Higgs hunter's guide*, volume 80, page 195. Brookhaven National Laboratory, 1989.

[11] G. Aad et al. [The ATLAS Collaboration]. Searches for Higgs boson pair production in the $hh \to bb\tau\tau$, $\gamma\gamma WW^*$, $\gamma\gamma bb$, $bbbb$ channels with the ATLAS detector. *Physical Review D*, 92(9):092004, 2015.

[12] G. Aad et al. [The ATLAS Collaboration]. Search for Higgs boson pair production in the $b\bar{b}b\bar{b}$ final state from pp collisions at $\sqrt{s} = 8\,\text{TeV}$ with the ATLAS detector. *The European Physical Journal C*, 75(9):412, 2015.

[13] G. Aad et al. [The ATLAS Collaboration]. Search for Higgs Boson Pair Production in the $\gamma\gamma b\bar{b}$ Final State Using pp Collision Data at $\sqrt{s} = 8\,\text{TeV}$ from the ATLAS Detector. *Phys. Rev. Lett.*, 114:081802, 2015.

[14] M. Sirunyan et al. [The CMS Collaboration]. Search for Higgs boson pair production in events with two bottom quarks and two tau leptons in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$. *Physics Letters B*, 778:101–127, 2018.

[15] F. Halzen and A. D. Martin. *Quarks and Leptons: An Introductory Course in Modern Particle Physics*. Wiley, 1984.

[16] M. E. Peskin and D. V. Schroeder. *An Introduction to Quantum Field Theory*, page 79. Perseus Books, 1995.

[17] P. A. M. Dirac. On the theory of quantum mechanics. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 112(762):661–677, 1926.

[18] W. Pauli. The Connection Between Spin and Statistics. *Physics Review*, 58:716–722, 1940.

[19] C. Patrignani et at. [Particle Data Group]. Review of Particle Physics. *Chinese Physics*, C40(10), 2016.

[20] M. Gell-Mann. Symmetries of Baryons and Mesons. *Phys. Rev.*, 125:1067–1084, 1962.

[21] S. L. Glashow. The renormalizability of vector meson interactions. *Nuclear Physics*, 10:107–117, 1959.

[22] A. Salam and J. C. Ward. Weak and electromagnetic interactions. *Il Nuovo Cimento (1955-1965)*, 11(4):568–577, 1959.

[23] S. Weinberg. A model of leptons. *Physical Review Letters*, 19:1264–1266, 1967.

[24] M. Bona et al. [UTfit Collaboration]. The Unitarity Triangle Fit in the Standard Model and Hadronic Parameters from Lattice QCD: A Reappraisal after the Measurements of Δm_s and $BR(B \to \tau\nu)$. *Journal of High Energy Physics*, 10:81, 2006.

[25] F. Englert and R. Brout. Broken Symmetry and the Mass of Gauge Vector Mesons. *Physical Review Letters*, 13:321–323, 1964.

[26] P.W. Higgs. Broken symmetries, massless particles and gauge fields. *Physics Letters*, 12(2):132–133, 1964.

[27] P. W. Higgs. Broken Symmetries and the Masses of Gauge Bosons. *Physical Review Letters*, 13:508–509, 1964.

[28] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. Global Conservation Laws and Massless Particles. *Phys. Rev. Lett.*, 13:585–587, 1964.

[29] J. Goldstone. Field theories with Superconductor solutions. *Il Nuovo Cimento (1955-1965)*, 19(1):154–164, 1961.

[30] M. Aaboud et al. [The ATLAS Collaboration]. Measurements of $t\bar{t}$ differential cross-sections of highly boosted top quarks decaying to all-hadronic final states in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector. Measurements of $t\bar{t}$ differential cross-sections of highly boosted top quarks decaying to all-hadronic final states in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector. *CERN-EP-2017-226*, 2018.

[31] G. Aad et al. [The ATLAS Collaboration]. Measurement of differential production cross-sections for a Z boson in association with b-jets in 7 TeV proton-proton collisions with the ATLAS detector. *Journal of High Energy Physics*, 2014(10), 2014.

[32] J. Neyman and E. S. Pearson. On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London A*, 231(694-706):289–337, 1933.

[33] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, and H. Voss. TMVA: Toolkit for Multivariate Data Analysis. *PoS*, ACAT:40, 2007.

[34] N. Cristianini and J. Shawe-Taylor. *An Introduction to Support Vector Machines.* Cambridge University Press, 2000.

[35] C. M. Bishop. *Pattern Recognition and Machine Learning.* Springer, 2006.

[36] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning.* Springer, 2009.

[37] L. Breiman. Bagging Predictors. *Machine Learning*, 24:123–140, 1996.

[38] Y. Freund and R. E. Schapire. Experiments with a New Boosting Algorithm. *Proceedings of the Thirteenth International Conference on Machine Learning*, pages 148–156, 1996.

[39] C. Cortes and V. Vapnik. Support-Vector Networks. *Machine Learning*, 20(3):273–297, 1995.

[40] R. A. Fisher. The Use of Multiple Measurements in Taxonomic Problems. *Annals of Human Genetics*, 7(2):179–188, 1936.

[41] W. Karush. *Minima of Functions of Several Variables with Inequalities as Side Conditions*, pages 217–245. Springer Basel, 2014.

[42] H.W. Kuhn and A.W. Tucker. Nonlinear Programming. *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pages 481–492, 1951.

[43] J. Mercer. Functions of Positive and Negative Type, and their Connections with the Theory of Integral Equations. *Philosophical Transactions of the Royal Society*, A(209):441–458, 1909.

[44] S. Arlot and A. Celisse. A survey of cross-validation procedures for model selection. *Statistics Surveys*, 4:40–79, 2010.

[45] C. Lefévre. The CERN accelerator complex. *CERN-DI-0812015*, 2008.

[46] G. Aad et al. [The ATLAS Collaboration]. The ATLAS Experiment at the CERN Large Hadron Collider. *Journal of Instrumentation*, 3(S08003), 2008.

[47] G. Aad et al. [The ATLAS Collaboration]. The ATLAS Inner Detector commissioning and calibrations. *The European Physics Journal C*, 70(787), 2010.

[48] A. La Rosa and on behalf of ATLAS collaboration. The ATLAS Insertable B-Layer: from construction to operation. *Journal of Instrumentation*, 11(12):C12036, 2016.

[49] M. Capeans, G. Darbo, K. Einsweiller, M. Elsing, T. Flick, M. Garcia-Sciveres, C. Gemme, H. Pernegger, O. Rohne, and R. Vuillermet. ATLAS Insertable B-Layer Technical Design Report. *CERN-LHCC-2010-013*, 2010.

[50] S. Agostinelli et al. [GEANT4 collaboration]. Geant4 – a simulation toolkit. *Nuclear Instruments and Methods in Physics Research*, A506(3):250–303, 2003.

[51] G. Aad et al. [The ATLAS Collaboration]. The ATLAS Simulation Infrastructure. *The European Physics Journal C*, 70(3):823–874, 2010.

[52] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer. MadGraph 5: going beyond. *Journal of High Energy Physics*, 2011(6):128, 2011.

[53] T. Sjöstrand, S. Mrenna, and P. Skands. A brief introduction to PYTHIA 8.1. *Computer Physics Communications*, 178(11):852–867, 2008.

[54] G. Aad et al. [The ATLAS Collaboration]. ATLAS Run 1 Pythia8 tunes. *ATL-PHYS-PUB-2014-021*, 2014.

[55] R. D. Ball, V. Bertone, S. Carrazza, C. S. Deans, L. Del Debbio, S. Forte, A. Guffanti, N. P. Hartland, J. I. Latorre, J. Rojo, and M. Ubiali. Parton distributions with LHC data. *Nuclear Physics B*, 867(2):244–289, 2013.

[56] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *Journal of High Energy Physics*, 2014(7):79, 2014.

[57] M. Bähr, S. Gieseke, M. A. Gigg, D. Grellscheid, K. Hamilton, O. Latunde-Dada, S. Plätzer, P. Richardson, M. H. Seymour, A. Sherstnev, and B. R. Webber. Herwig++ physics and manual. *The European Physical Journal C*, 58(4):639–707, 2008.

[58] J. Pumplin, D. R. Stump, J. Huston, H. L. Lai, P. Nadolsky, and W. K. Tung. New Generation of Parton Distributions with Uncertainties from Global QCD Analysis. *Journal of High Energy Physics*, 2002(7):12, 2002.

[59] S. Alioli, P. Nason, C. Oleari, and E. Re. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *Journal of High Energy Physics*, 2010(6):43, 2010.

[60] T. Ježo, J. M. Lindert, P. Nason, C. Oleari, and S.Pozzorini. An NLO+PS generator for $t\bar{t}$ and Wt production and decay including non-resonant and interference effects. *The European Physical Journal C*, 76(12):691, 2016.

[61] S. Alioli, P. Nason, C. Oleari, and E. Re. Erratum: NLO single-top production matched with shower in POWHEG: s- and t-channel contributions. *Journal of High Energy Physics*, 2010(2):11, 2010.

[62] H. L. Lai, M. Guzzi, J. Huston, Z. Li, P. M. Nadolsky, J. Pumplin, and C. P. Yuan. New parton distributions for collider physics. *Phys. Rev. D*, 82:074024, 2010.

[63] P. Artoisenet, R. Frederix, O. Mattelaer, and Robbert R. Rietkerk. Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations. *Journal of High Energy Physics*, 2013(3):15, 2013.

[64] T. Sjöstrand, S. Mrenna, and P. Skands. PYTHIA 6.4 physics and manual. *Journal of High Energy Physics*, 2006(5):26, 2006.

[65] P. Skands. Tuning Monte Carlo generators: The Perugia tunes. *Physical Review D*, 82:074018, 2010.

[66] D. J. Lange. The EvtGen particle decay simulation package. *Nuclear Instruments and Methods in Physics Research*, 462(1):152–155, 2001.

[67] M. Czakon, M. L. Mangano, A. Mitov, and J. Rojo. Constraints on the gluon PDF from top quark pair production at hadron colliders. *Journal of High Energy Physics*, 2013(7):167, 2013.

[68] T. Gleisberg, S. Höche, F. Krauss, M. Schönherr, S. Schumann, F. Siegert, and J. Winter. Event generation with SHERPA 1.1. *Journal of High Energy Physics*, 2009(2):7, 2009.

[69] T. Gleisberg and S. Höche. Comix, a new matrix element generator. *Journal of High Energy Physics*, 2008(12):039, 2008.

[70] F. Cascioli, P. Maierhöfer, and S. Pozzorini. Scattering Amplitudes with Open Loops. *Physical Review Letters*, 108:111601, 2012.

[71] S. Schumann and F. Krauss. A parton shower algorithm based on Catani-Seymour dipole factorisation. *Journal of High Energy Physics*, 2008(03):038, 2008.

[72] S. Höche, F. Krauss, M. Schönherr, and F. Siegert. QCD matrix elements + parton showers. The NLO case. *Journal of High Energy Physics*, 2013(4):27, 2013.

[73] S. Alioli, P. Nason, C. Oleari, and E. Re. NLO Higgs boson production via gluon fusion matched with shower in POWHEG. *Journal of High Energy Physics*, 2009(4):2, 2009.

[74] G. Aad et al. [The ATLAS Collaboration]. Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s} = 7\,\text{TeV}$ with the ATLAS detector. *Journal of High Energy Physics*, 2014(9):145, 2014.

[75] M. Cacciari, G. Salam, and G. Soyez. The anti-k_t jet clustering algorithm. *Journal of High Energy Physics*, 0804:63, 2008.

[76] G. Aad et al. [The ATLAS Collaboration]. Jet energy measurement and its systematic uncertainty in proton–proton collisions at $\sqrt{s} = 7\,\text{TeV}$ with the ATLAS detector. *European Physical Journal*, C 75:17, 2105.

[77] G. Aad et al. [The ATLAS Collaboration]. Pile-up subtraction and suppression for jets in ATLAS. *ATLAS-CONF-2013-083*, 2013.

[78] G. Aad et al. [The ATLAS Collaboration]. Performance of b-jet identification in the ATLAS experiment. *Journal of Instrumentation*, 11(4):4008, 2016.

[79] G. Aad et al. [The ATLAS Collaboration]. Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run. *ATL-PHYS-PUB-2016-012*, 2016.

[80] G. Aad et al. [The ATLAS Collaboration]. Object selections for the SM Vh and $A \to Zh$ search with V decaying leptonically and $h \to b\bar{b}$ with Run-2 data collected with the ATLAS detector at $\sqrt{s} = 13\,\text{TeV}$. *ATL-COM-PHYS-2015-1467*, 2015.

[81] G. Aad et al. [The ATLAS Collaboration]. Electron reconstruction and identification efficiency measurements with the ATLAS detector using the 2011 LHC proton–proton collision data. *European Physical Journal*, C 74:2941, 2014.

[82] G. Aad et al. [The ATLAS Collaboration]. Muon reconstruction performance of the ATLAS detector in proton-proton collision data at $\sqrt{s} = 13\,\text{TeV}$. *European Physical Journal*, C 76:292, 2016.

[83] G. Aad et al. [The ATLAS Collaboration]. Identification and energy calibration of hadronically decaying tau leptons with the ATLAS experiment in pp collisions at $\sqrt{s} = 8\,\text{TeV}$. *European Physical Journal*, C 75:303, 2015.

[84] G. Aad et al. [The ATLAS Collaboration]. Reconstruction, Energy Calibration, and Identification of Hadronically Decaying Tau Leptons in the ATLAS Experiment for Run-2 of the LHC. *ATL-PHYS-PUB-2015-045*, 2015.

[85] G. Aad et al. [The ATLAS Collaboration]. Measurement of the tau lepton reconstruction and identification performance in the ATLAS experiment using pp collisions at $\sqrt{s} = 13\,\text{TeV}$. *ATLAS-CONF-2017-029*, 2017.

[86] G. Aad et al. [The ATLAS Collaboration]. Expected performance of missing transverse momentum reconstruction for the ATLAS detector at $\sqrt{s} = 13\,\text{TeV}$. *ATL-PHYS-PUB-2015-023*, 2015.

[87] A. Elagin, P. Murat, A. Pranko, and A. Safonov. A New Mass Reconstruction Technique for Resonances Decaying to $\tau\tau$. *Nuclear Instruments and Methods in Physics Research*, A654:481–489, 2011.

[88] G. Aad et al. [The ATLAS Collaboration]. Selection of jets produced in 13TeV proton-proton collisions with the ATLAS detector. *ATLAS-CONF-2015-029*, 2015.

[89] M. Aaboud et al. [The ATLAS Collaboration]. Luminosity determination in pp collisions at $\sqrt{s} = 8\,\text{TeV}$ using the ATLAS detector at the LHC. *The European Physical Journal C*, 76(12):653, 2016.

[90] M. Aaboud et al. [The ATLAS Collaboration]. Jet energy scale measurements and their systematic uncertainties in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector. *Physics Review D*, 96, 2017.

[91] G. Aad et al. [The ATLAS Collaboration]. Jet energy resolution in proton-proton collisions at $\sqrt{s} = 7\,\text{TeV}$ recorded in 2010 with the ATLAS detector. *The European Physical Journal C*, 73(3):2306, 2013.

[92] G. Aad et al. [The ATLAS Collaboration]. Jet Calibration and Systematic Uncertainties for Jets Reconstructed in the ATLAS Detector at $\sqrt{s} = 13\,\text{TeV}$. *ATL-PHYS-PUB-2015-015*, 2015.

[93] G. Aad et al. [The ATLAS Collaboration]. Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton–proton collision data. *ATLAS-CONF-2016-024*, 2016.

[94] J. Butterworth, S. Carrazza, A. Cooper-Sarkar, A. De Roeck, J. Feltesse, S. Forte, J. Gao, S. Glazov, J. Huston, Z. Kassabov, R. McNulty, A. Morsch, P. Nadolsky, V. Radescu, J. Rojo, and R. Thorne. PDF4LHC recommendations for LHC Run II. *Journal of Physics G: Nuclear and Particle Physics*, 43(2):023001, 2016.

[95] S. Borowka, N. Greiner, G. Heinrich, S. P. Jones, M. Kerner, J. Schlenk, U. Schubert, and T. Zirke. Higgs Boson Pair Production in Gluon Fusion at Next-to-Leading Order with Full Top-Quark Mass Dependence. *Physical Review Letters*, 117:012001, 2016.

[96] L. Moneta, K. Belasco, K. Cranmer, S. Kreiss, A. Lazzaro, D. Piparo, G. Schott, W. Verkerke, and M. Wolf. The RooStats Project. In *13th International Workshop on Advanced Computing and Analysis Techniques in Physics Research*. Proceedings of Science, 2010.

[97] A L Read. Presentation of search results: the CL s technique. *Journal of Physics G: Nuclear and Particle Physics*, 28(10):2693, 2002.

[98] G. Cowan, K. Cranmer, E. Gross, and O. Vitells. Asymptotic formulae for likelihood-based tests of new physics. *The European Physical Journal C*, 71(2):1554, 2011.

[99] A. Djouadi, L. Maiani, G. Moreau, A. Polosa, J. Quevillon, and V. Riquer. The post-Higgs MSSM scenario: habemus MSSM? *The European Physical Journal C*, 73(12):2650, 2013.

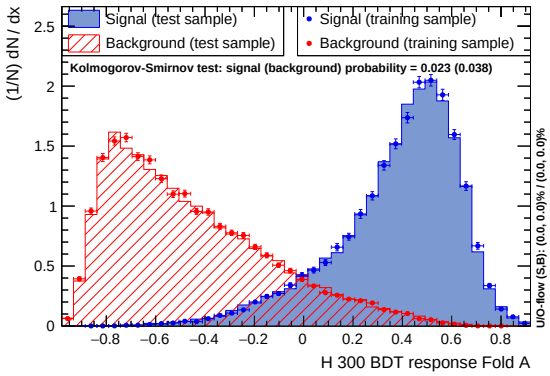
[100] F. James. MINUIT Function Minimization and Error Analysis: Reference Manual Version 94.1. *CERN-D-506*, 1994.

Appendix A

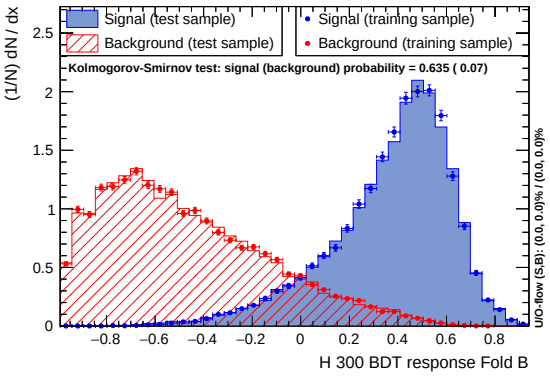
MVA Trainings

The figures shown in the following are control plots for the nominal BDT trainings used in the analysis, described in Section 7.3. For a representative selection of mass points they include the BDT training output distributions and the ROC curves as well as pre-fit data-MC comparisons of the BDT distributions:
Figure A.1 shows the BDT output distributions for the 2HDM signal in the SLT channel.
Figure A.2 shows the BDT output distributions for the RSG signal in the SLT channel.
Figure A.3 shows the BDT output distributions for the 2HDM signal in the LTT channel.
Figure A.4 shows the BDT output distributions for the RSG signal in the LTT channel.
Figure A.5 shows the BDT output distributions for the non-resonant SM signal in the SLT and LTT channels.
Figure A.6 shows ROC curves for the 2HDM signal in the SLT channel.
Figure A.7 shows ROC curves for the RSG signal in the SLT channel.
Figure A.8 shows ROC curves for the 2HDM signal in the LTT channel.
Figure A.9 shows ROC curves for the RSG signal in the LTT channel.
Figure A.10 shows ROC curves for the non-resonant SM signal in the SLT and LTT channels.
Figure A.11 shows the pre-fit data-MC comparisons of the BDT distributions for events at preselection with two $b-$tagged jets for the 2HDM signal in the SLT channel.
Figure A.12 shows the pre-fit data-MC comparisons of the BDT distributions for events at preselection with two $b-$tagged jets for the RSG signal in the SLT channel.
Figure A.13 shows the pre-fit data-MC comparisons of the BDT distributions for events at preselection with two $b-$tagged jets for the 2HDM signal in the LTT channel.
Figure A.14 shows the pre-fit data-MC comparisons of the BDT distributions for events at preselection with two $b-$tagged jets for the RSG signal in the LTT channel.
Figure A.15 shows the pre-fit data-MC comparisons of the BDT distributions for events at preselection with two $b-$tagged jets for the non-resonant SM signal in the SLT and LTT channels.

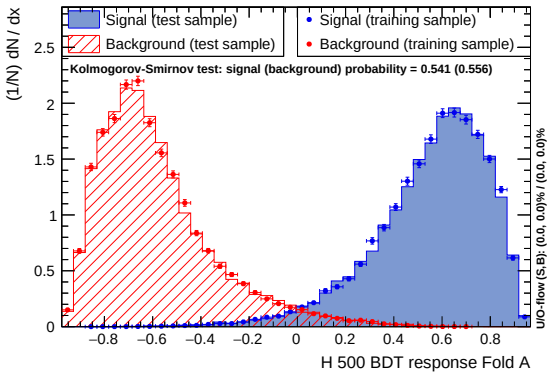
A.1 SLT 2HDM BDT Trainings



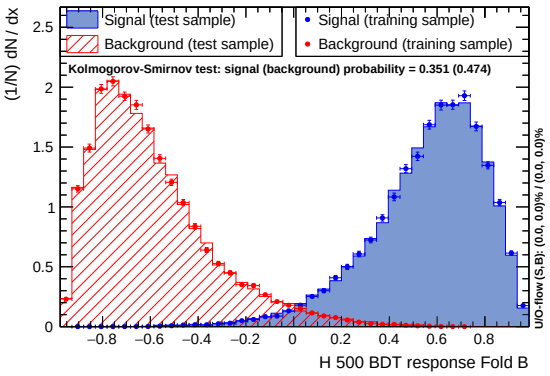
(a) 2HDM ($m_H = 300\,\text{GeV}$) Fold 1



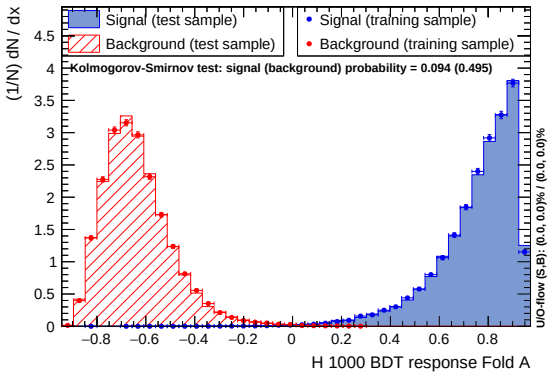
(b) 2HDM ($m_H = 300\,\text{GeV}$) Fold 2



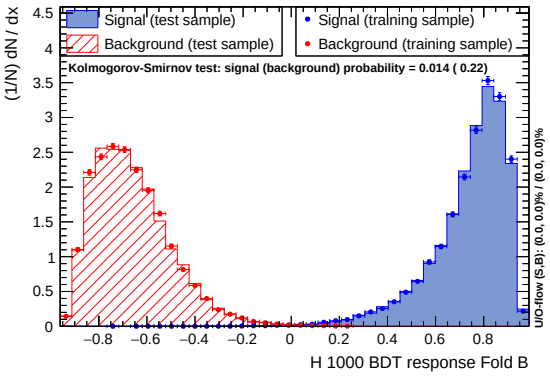
(c) 2HDM ($m_H = 500\,\text{GeV}$) Fold 1



(d) 2HDM ($m_H = 500\,\text{GeV}$) Fold 2



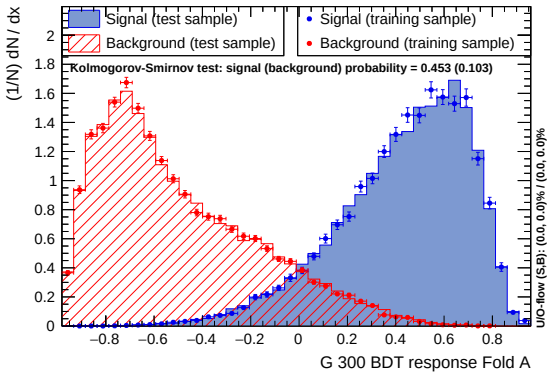
(e) 2HDM ($m_H = 1000\,\text{GeV}$) Fold 1



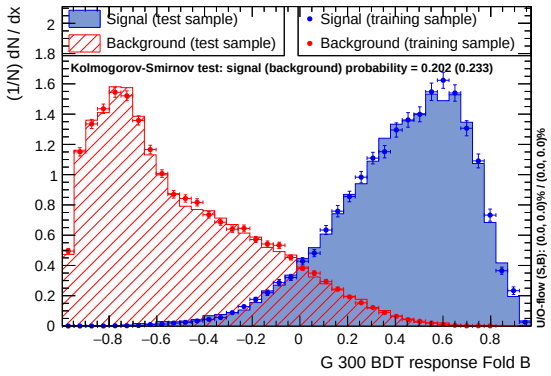
(f) 2HDM ($m_H = 1000\,\text{GeV}$) Fold 2

Figure A.1: BDT output distributions for the BDT folds trained on the $t\bar{t}$ background and the 2HDM signal for the 300 GeV (top), 500 GeV (middle) and 1000 GeV (bottom) mass points in the SLT channel.

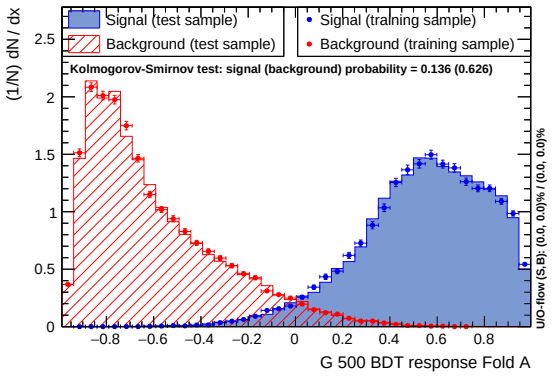
A.2 SLT RSG BDT Trainings



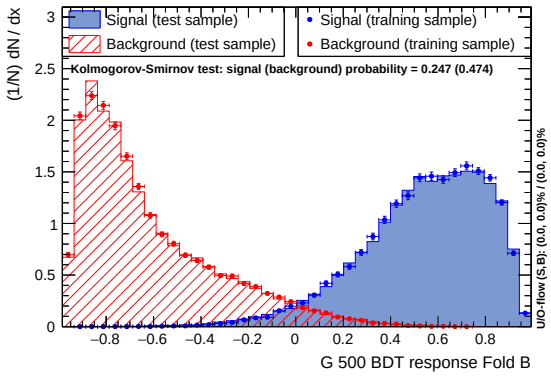
(a) RSG ($m_G = 300\,\text{GeV}$) Fold 1



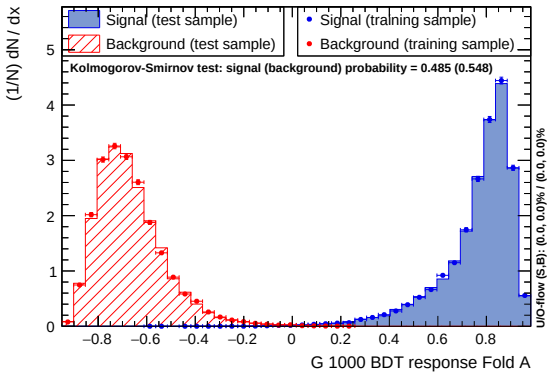
(b) RSG ($m_G = 300\,\text{GeV}$) Fold 2



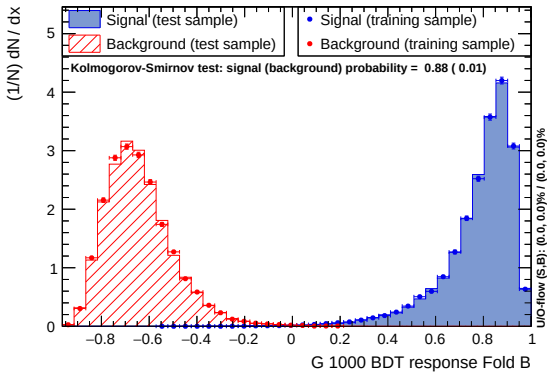
(c) RSG ($m_G = 500\,\text{GeV}$) Fold 1



(d) RSG ($m_G = 500\,\text{GeV}$) Fold 2



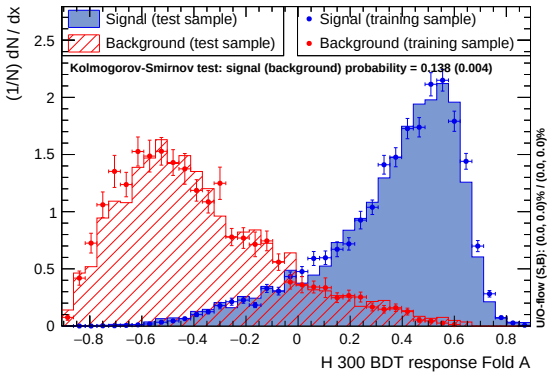
(e) RSG ($m_G = 1000\,\text{GeV}$) Fold 1



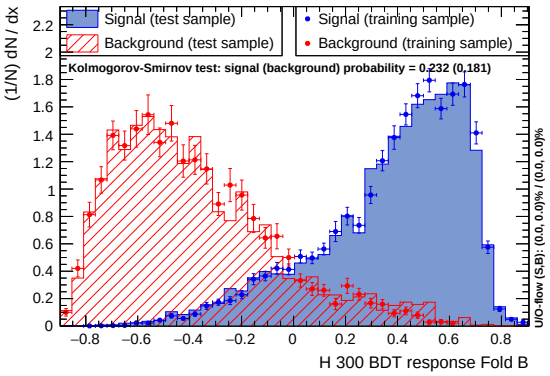
(f) RSG ($m_G = 1000\,\text{GeV}$) Fold 2

Figure A.2: BDT output distributions for the BDT folds trained on the $t\bar{t}$ background and the RSG signal for the 300 GeV (top), 500 GeV (middle) and 1000 GeV (bottom) mass points in the SLT channel.

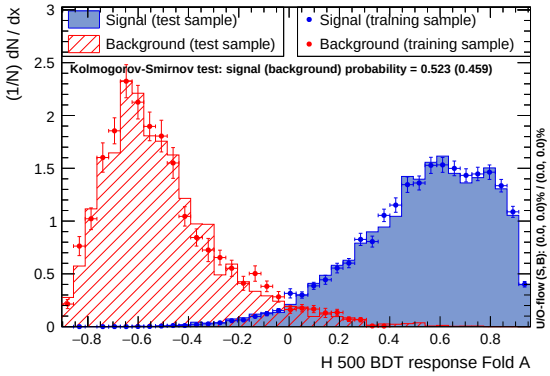
A.3 LTT 2HDM BDT Trainings



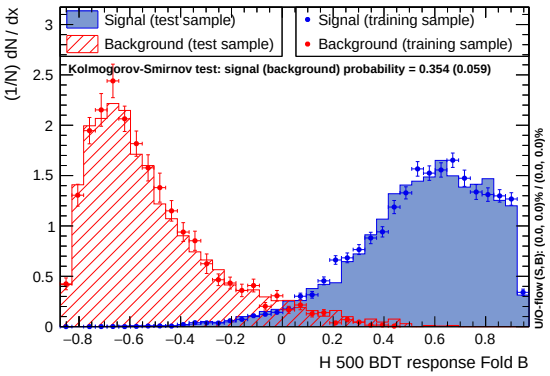
(a) 2HDM ($m_H = 300\,\text{GeV}$) Fold 1



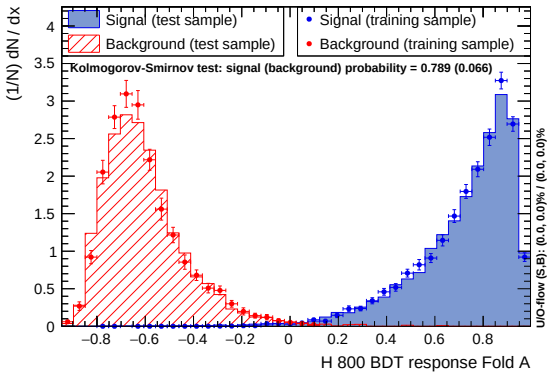
(b) 2HDM ($m_H = 300\,\text{GeV}$) Fold 2



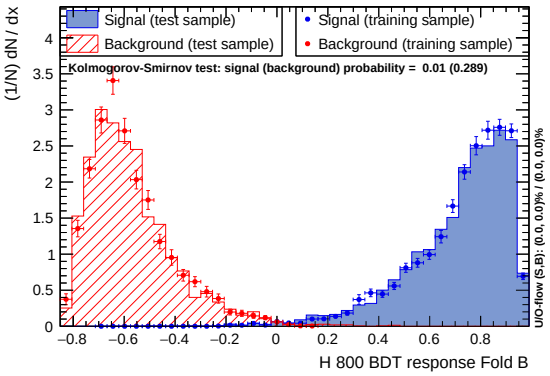
(c) 2HDM ($m_H = 500\,\text{GeV}$) Fold 1



(d) 2HDM ($m_H = 500\,\text{GeV}$) Fold 2



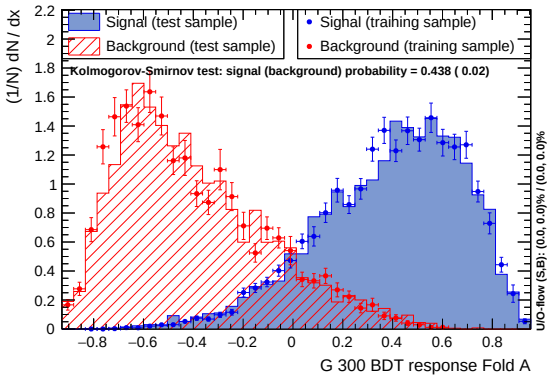
(e) 2HDM ($m_H = 800\,\text{GeV}$) Fold 1



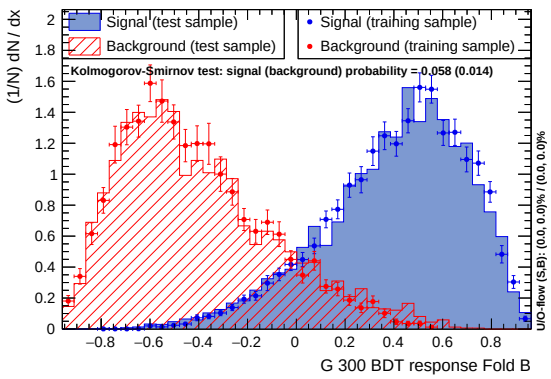
(f) 2HDM ($m_H = 800\,\text{GeV}$) Fold 2

Figure A.3: BDT output distributions for the BDT folds trained on the $t\bar{t}$ background and the 2HDM signal for the 300 GeV (top), 500 GeV (middle) and 800 GeV (bottom) mass points in the LTT channel.

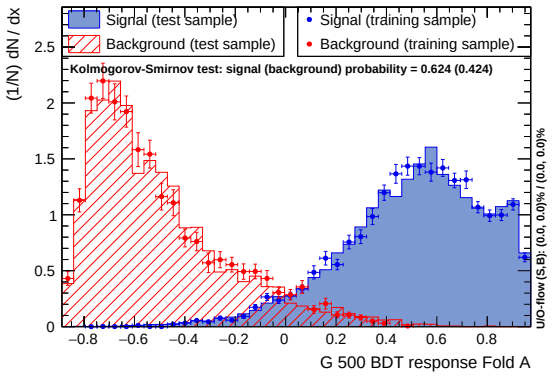
A.4 LTT RSG BDT Trainings



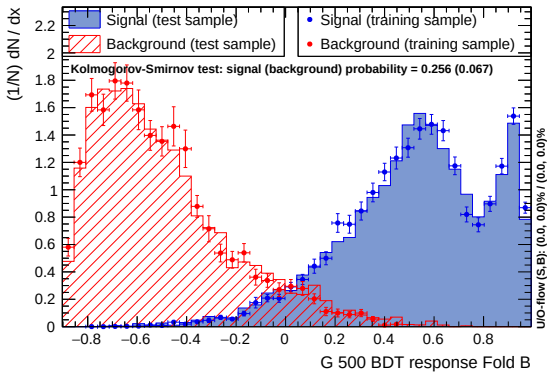
(a) RSG ($m_G = 300\,\text{GeV}$) Fold 1



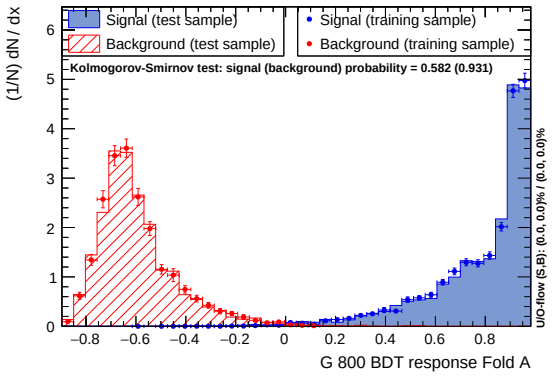
(b) RSG ($m_G = 300\,\text{GeV}$) Fold 2



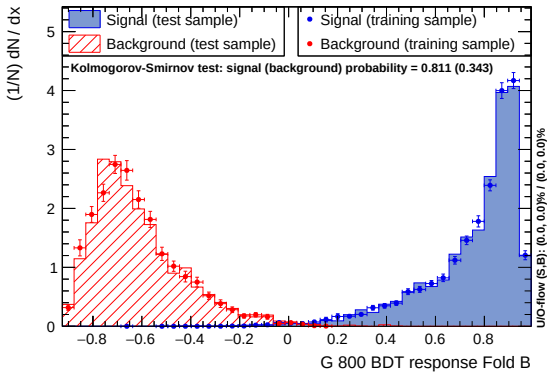
(c) RSG ($m_G = 500\,\text{GeV}$) Fold 1



(d) RSG ($m_G = 500\,\text{GeV}$) Fold 2



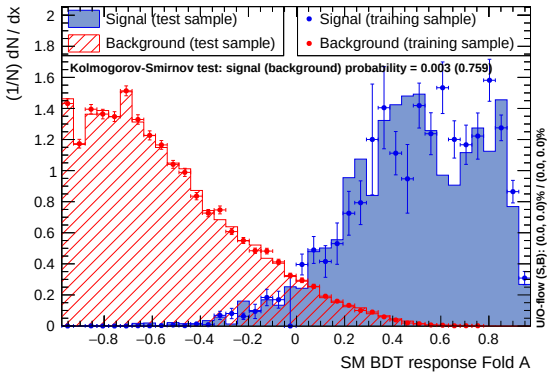
(e) RSG ($m_G = 800\,\text{GeV}$) Fold 1



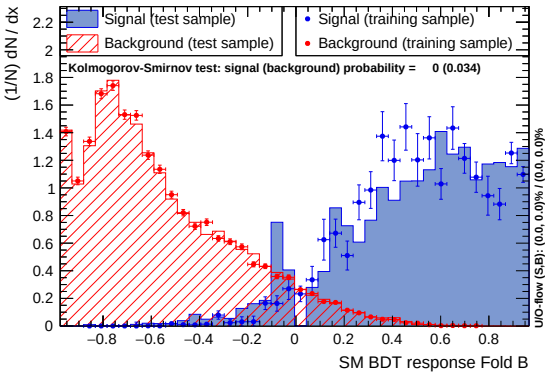
(f) RSG ($m_G = 800\,\text{GeV}$) Fold 2

Figure A.4: BDT output distributions for the BDT folds trained on the $t\bar{t}$ background and the RSG signal for the 300 GeV (top), 500 GeV (middle) and 800 GeV (bottom) mass points in the LTT channel.

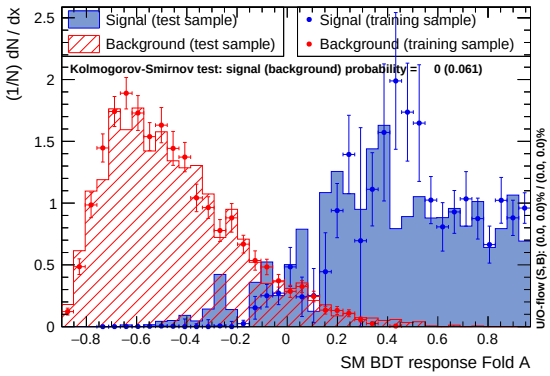
A.5 SM BDT Trainings



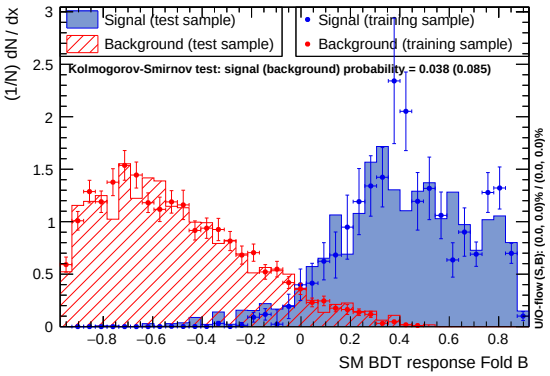
(a) SLT Fold 1



(b) SLT Fold 2



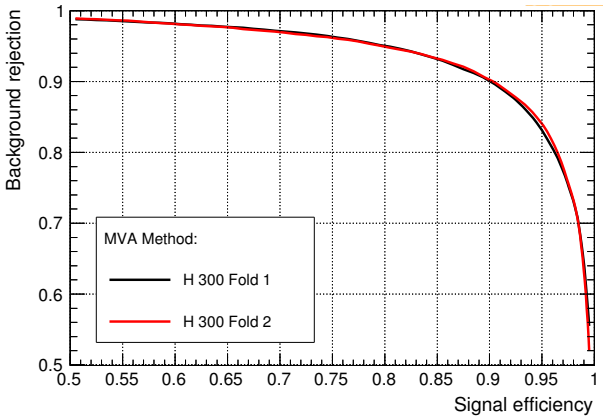
(c) SM Fold 1



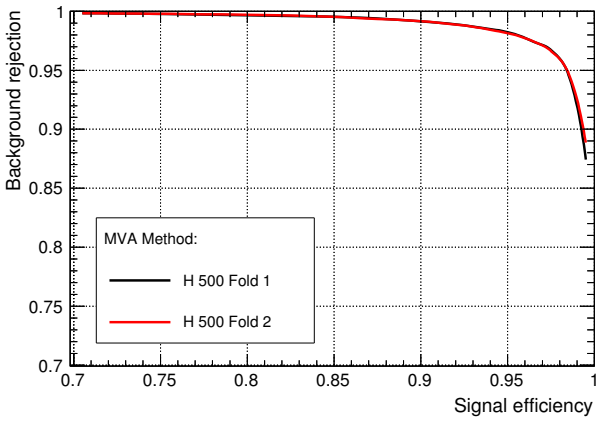
(d) SM Fold 2

Figure A.5: BDT output distributions for the BDT folds trained on the $t\bar{t}$ background and the SM non-resonant signal for the SLT channel (top) and LTT channel (bottom).

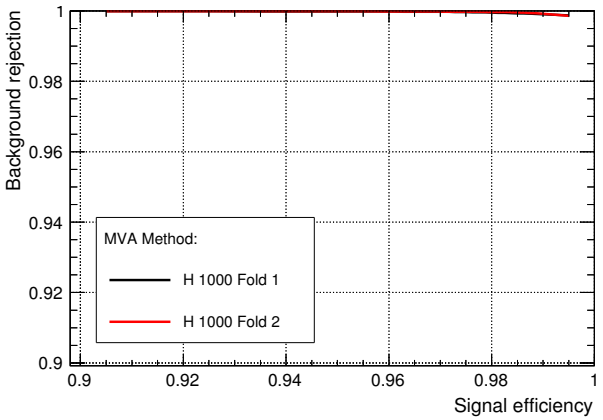
A.6 SLT 2HDM ROC Curves



(a) 2HDM ($m_H = 300\,\text{GeV}$)



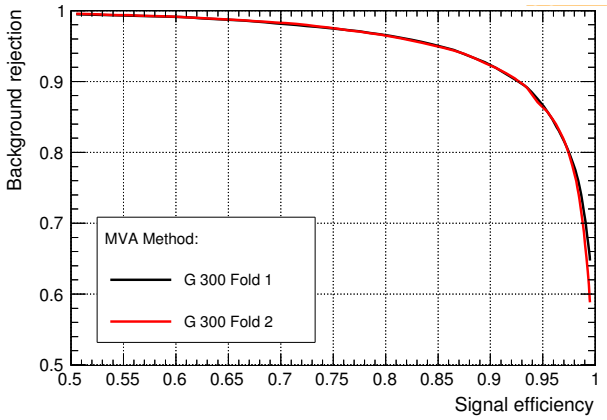
(b) 2HDM ($m_H = 500\,\text{GeV}$)



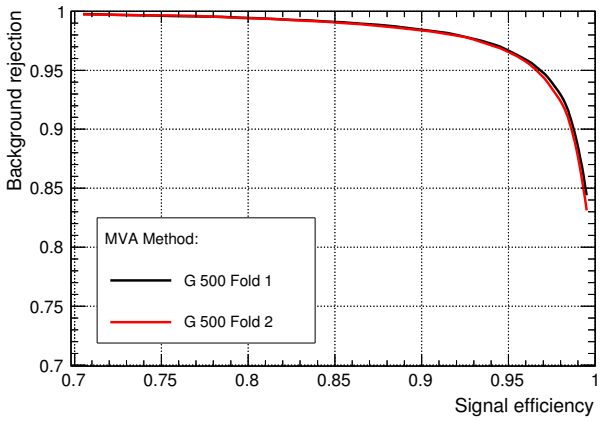
(c) 2HDM ($m_H = 1000\,\text{GeV}$)

Figure A.6: ROC curves comparing the BDT folds trained on the $t\bar{t}$ background and the 2HDM signal for the (a) 300 GeV, (b) 500 GeV and (c) 1000 GeV mass points in the SLT channel. The separation between signal and background increases significantly at higher mass points as the di-Higgs invariant mass variable dominates.

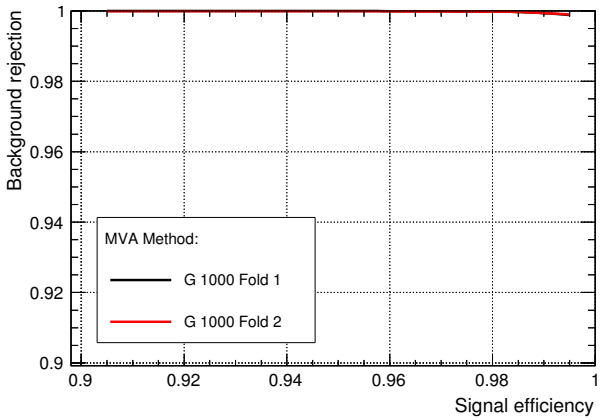
A.7 SLT RSG ROC Curves



(a) RSG ($m_G = 300\,\text{GeV}$)



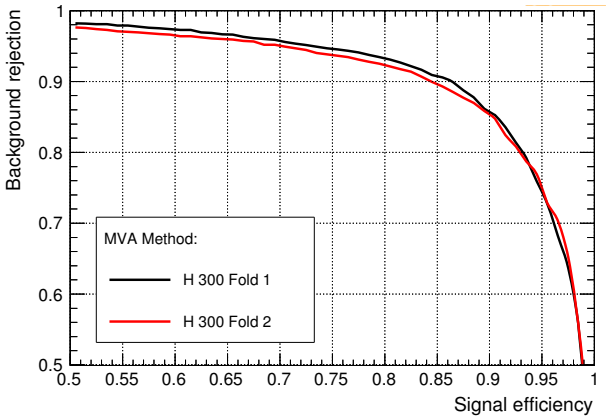
(b) RSG ($m_G = 500\,\text{GeV}$)



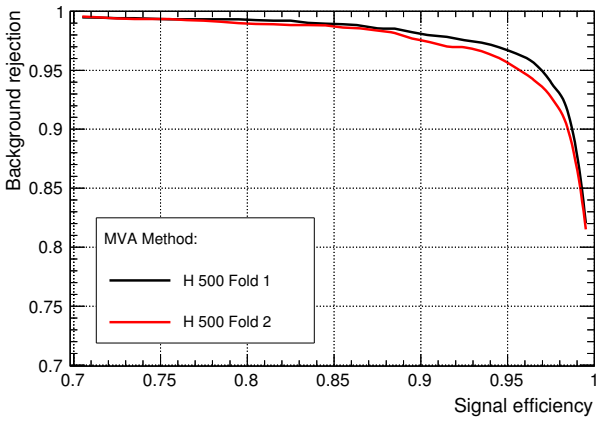
(c) RSG ($m_G = 1000\,\text{GeV}$)

Figure A.7: ROC curves comparing the BDT folds trained on the $t\bar{t}$ background and the RSG signal for the (a) 300 GeV, (b) 500 GeV and (c) 1000 GeV mass points in the SLT channel. The separation between signal and background increases significantly at higher mass points as the di-Higgs invariant mass variable dominates.

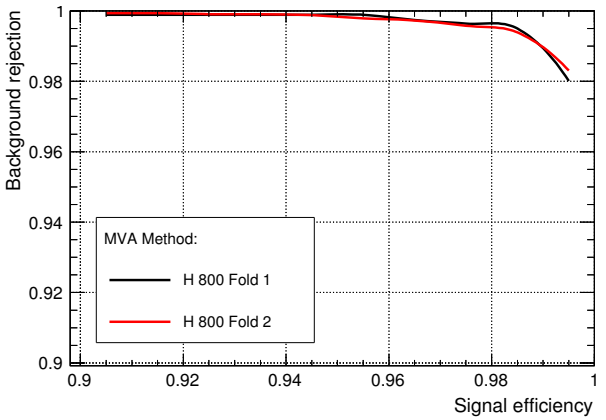
A.8 LTT 2HDM ROC Curves



(a) 2HDM ($m_H = 300\,\text{GeV}$)



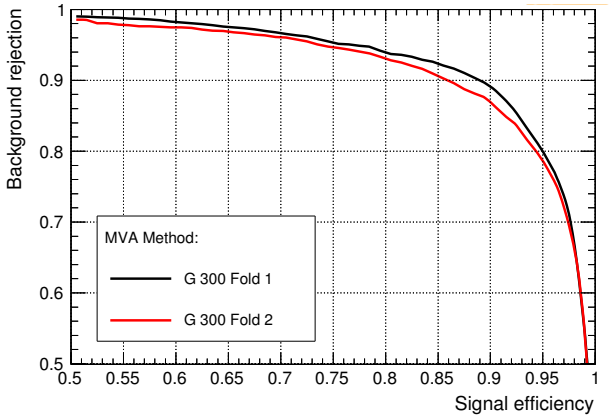
(b) 2HDM ($m_H = 500\,\text{GeV}$)



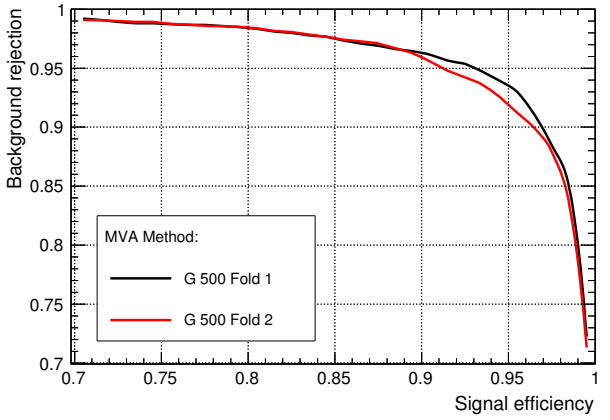
(c) 2HDM ($m_H = 800\,\text{GeV}$)

Figure A.8: ROC curves comparing the BDT folds trained on the $t\bar{t}$ background and the 2HDM signal for the (a) 300 GeV, (b) 500 GeV and (c) 800 GeV mass points in the LTT channel. The separation between signal and background increases significantly at higher mass points as the di-Higgs invariant mass variable dominates.

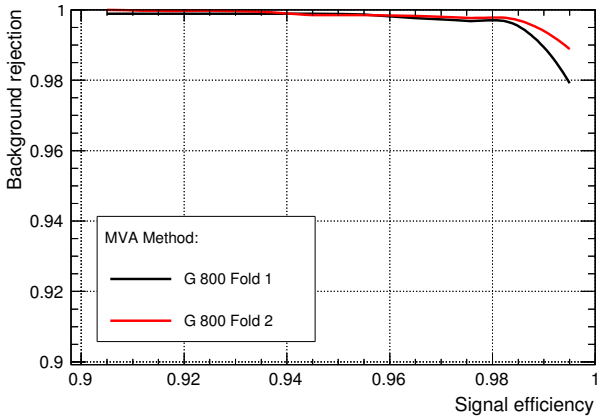
A.9 LTT RSG ROC Curves



(a) RSG ($m_G = 300\,\text{GeV}$)



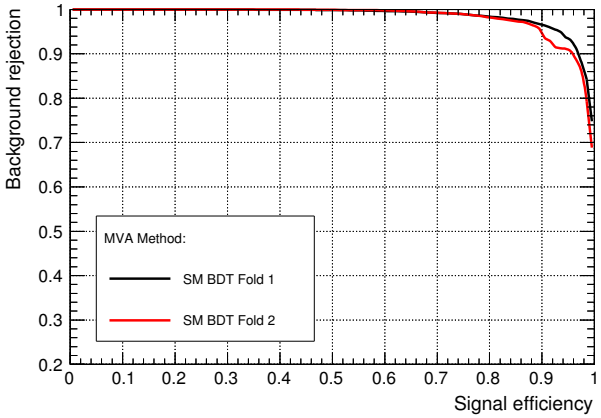
(b) RSG ($m_G = 500\,\text{GeV}$)



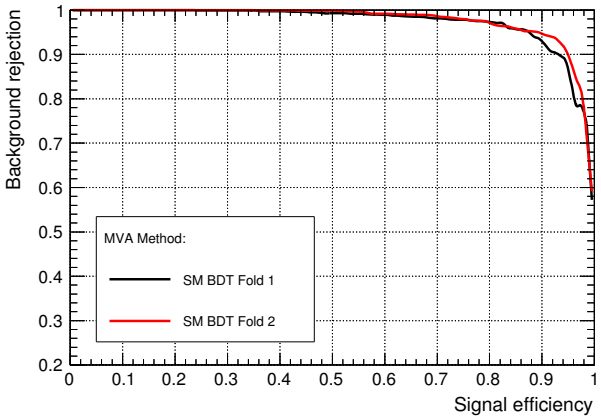
(c) RSG ($m_G = 800\,\text{GeV}$)

Figure A.9: ROC curves comparing the BDT folds trained on the $t\bar{t}$ background and the RSG signal for the (a) 300 GeV, (b) 500 GeV and (c) 800 GeV mass points in the LTT channel. The separation between signal and background increases significantly at higher mass points as the di-Higgs invariant mass variable dominates.

A.10 SM ROC Curves



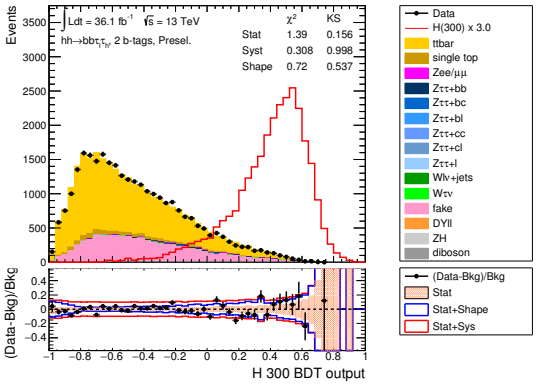
(a) SLT



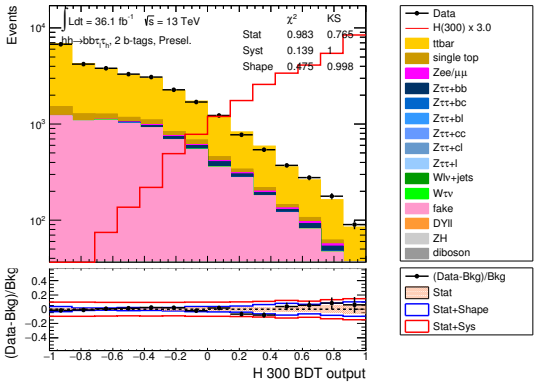
(b) LTT

Figure A.10: ROC curves comparing the BDT folds trained on the $t\bar{t}$ background and the SM signal for the (a) SLT and (b) LTT channels.

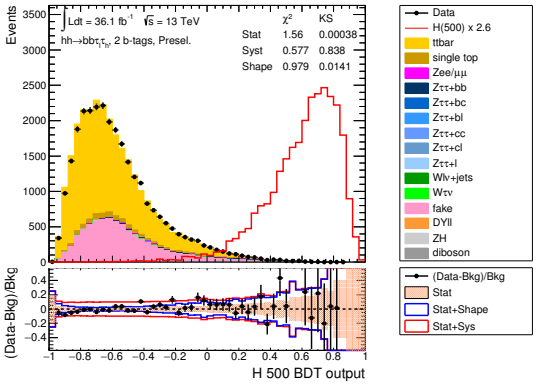
A.11 SLT 2HDM Prefit



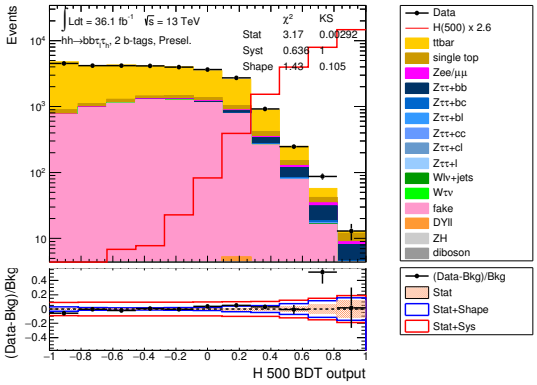
(a) 2HDM ($m_H = 300\,\text{GeV}$)



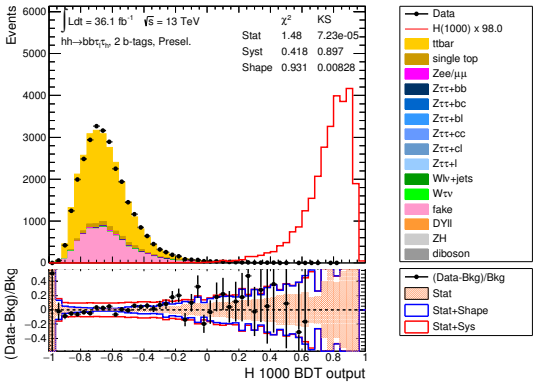
(b) 2HDM ($m_H = 300\,\text{GeV}$)



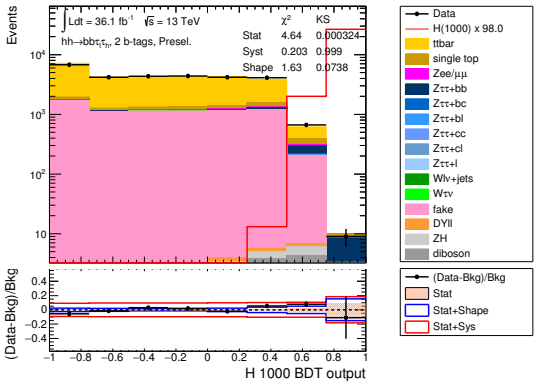
(c) 2HDM ($m_H = 500\,\text{GeV}$)



(d) 2HDM ($m_H = 500\,\text{GeV}$)



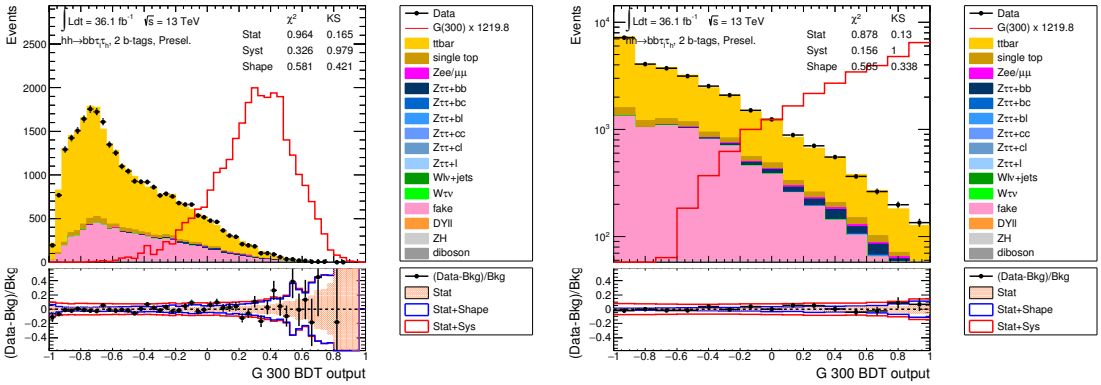
(e) 2HDM ($m_H = 1000\,\text{GeV}$)



(f) 2HDM ($m_H = 1000\,\text{GeV}$)

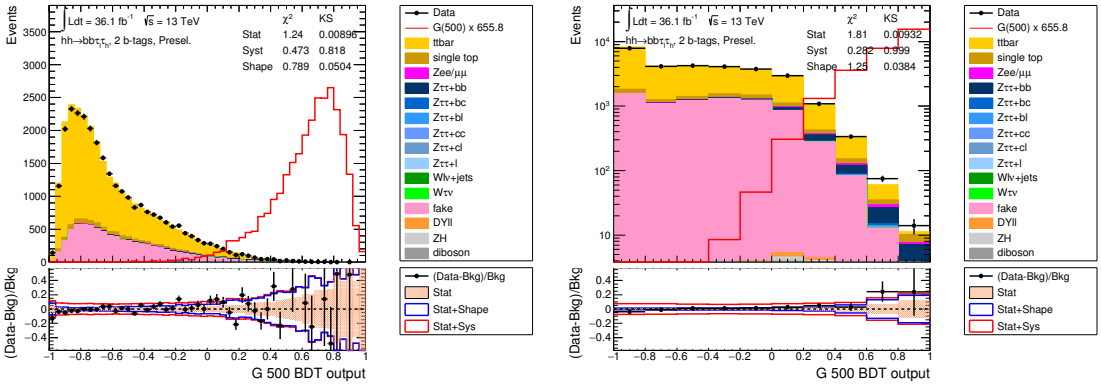
Figure A.11: Pre-fit data-MC comparisons of the BDT distributions for events at preselection with two $b-$tagged jets using arbitrary binning (left) and optimised binning (right). The 2HDM signal and all backgrounds for the (a) 300 GeV, (b) 500 GeV and (c) 1000 GeV mass points in the SLT channel are shown.

A.12 SLT RSG Prefit



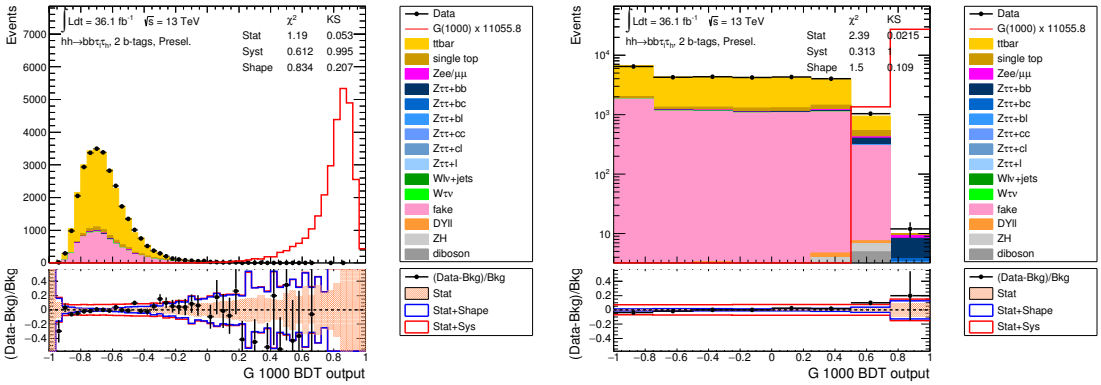
(a) RSG ($m_G = 300\,\text{GeV}$)

(b) RSG ($m_G = 300\,\text{GeV}$)



(c) RSG ($m_G = 500\,\text{GeV}$)

(d) RSG ($m_G = 500\,\text{GeV}$)

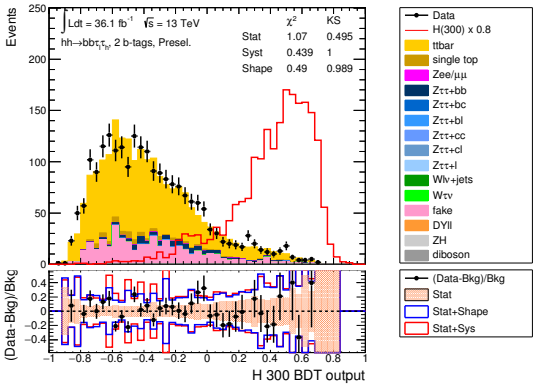


(e) RSG ($m_G = 1000\,\text{GeV}$)

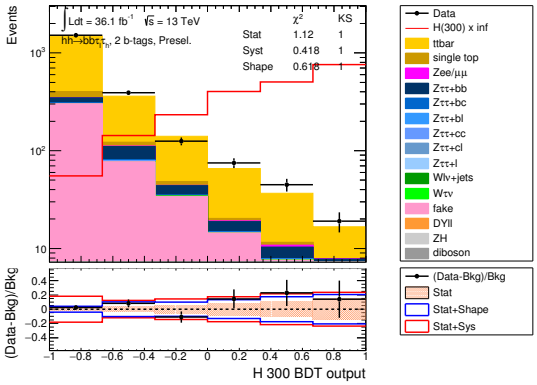
(f) RSG ($m_G = 1000\,\text{GeV}$)

Figure A.12: Pre-fit data-MC comparisons of the BDT distributions for events at preselection with two $b-$tagged jets using arbitrary binning (left) and optimised binning (right). The RSG signal and all backgrounds for the (a) 300 GeV, (b) 500 GeV and (c) 1000 GeV mass points in the SLT channel are shown.

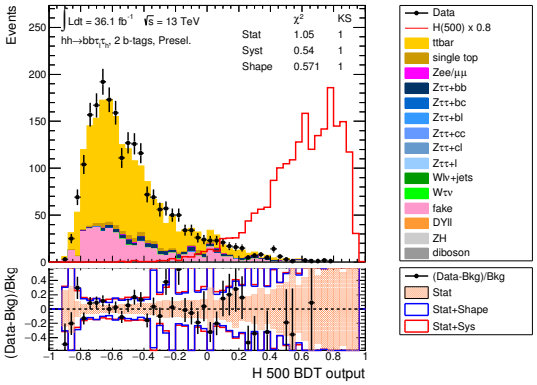
A.13 LTT 2HDM Prefit



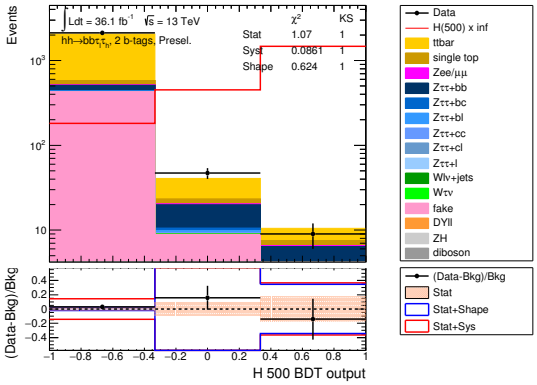
(a) 2HDM ($m_H = 300\,\text{GeV}$)



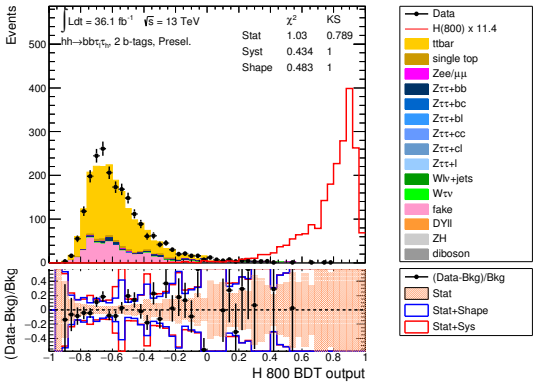
(b) 2HDM ($m_H = 300\,\text{GeV}$)



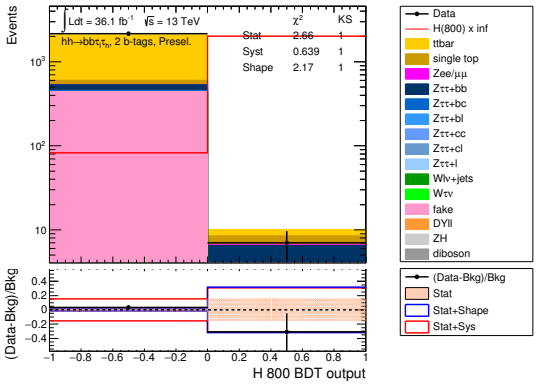
(c) 2HDM ($m_H = 500\,\text{GeV}$)



(d) 2HDM ($m_H = 500\,\text{GeV}$)



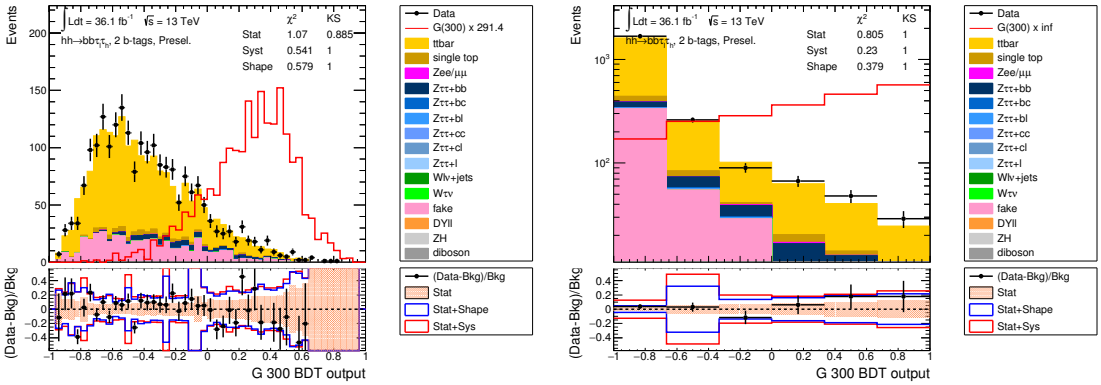
(e) 2HDM ($m_H = 1000\,\text{GeV}$)



(f) 2HDM ($m_H = 1000\,\text{GeV}$)

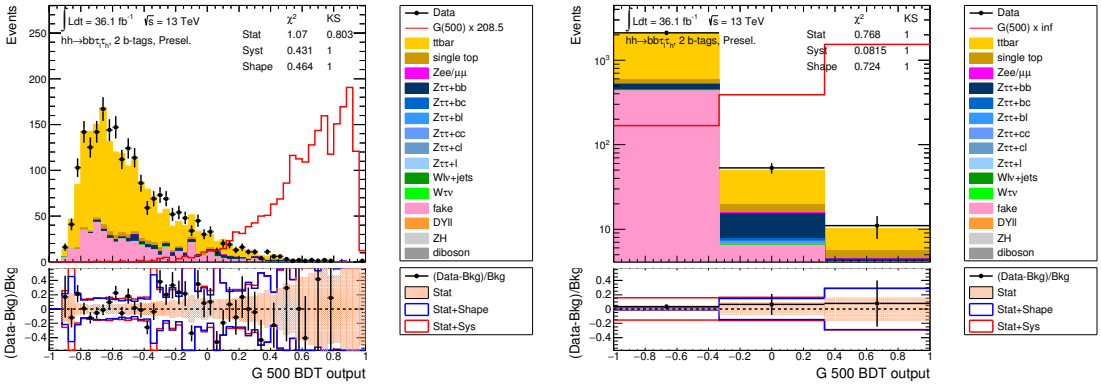
Figure A.13: Pre-fit data-MC comparisons of the BDT distributions for events at preselection with two $b-$tagged jets using arbitrary binning (left) and optimised binning (right). The 2HDM signal and all backgrounds for the (a) 300 GeV, (b) 500 GeV and (c) 800 GeV mass points in the LTT channel are shown.

A.14 LTT RSG Prefit



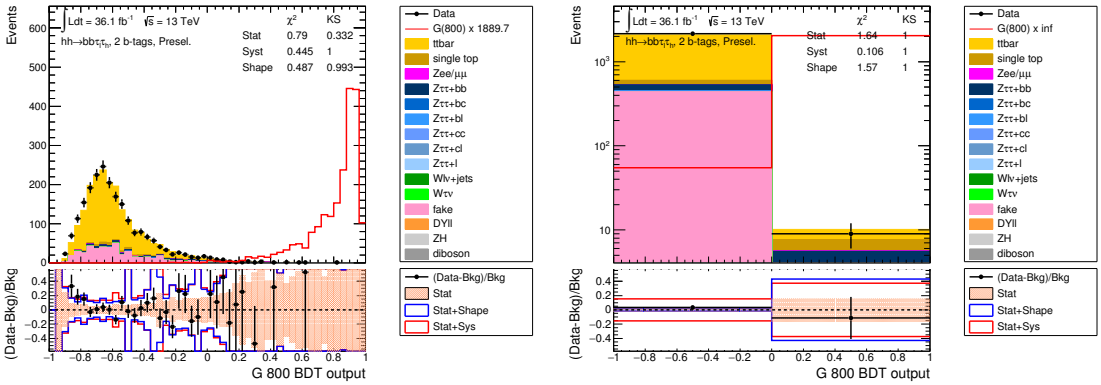
(a) RSG ($m_G = 300\,\text{GeV}$)

(b) RSG ($m_G = 300\,\text{GeV}$)



(c) RSG ($m_G = 500\,\text{GeV}$)

(d) RSG ($m_G = 500\,\text{GeV}$)

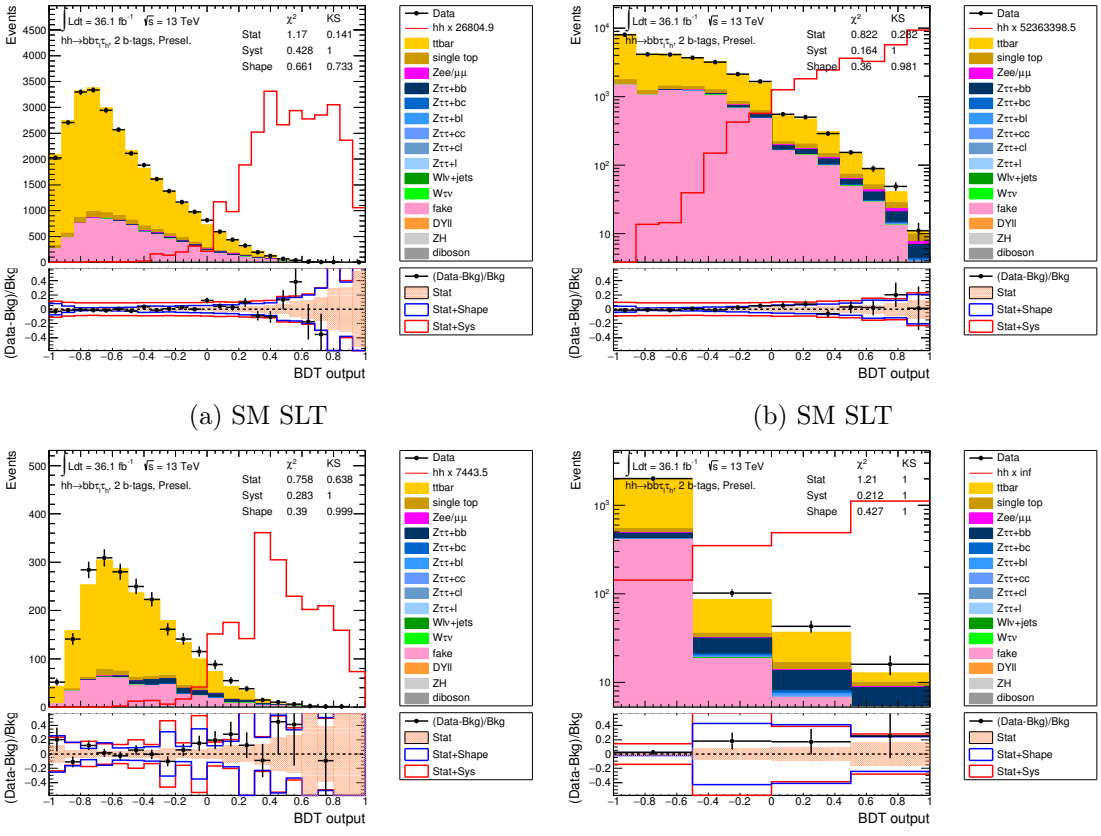


(e) RSG ($m_G = 1000\,\text{GeV}$)

(f) RSG ($m_G = 1000\,\text{GeV}$)

Figure A.14: Pre-fit data-MC comparisons of the BDT distributions for events at preselection with two $b-$tagged jets using arbitrary binning (left) and optimised binning (right). The RSG signal and all backgrounds for the (a) 300 GeV, (b) 500 GeV and (c) 800 GeV mass points in the LTT channel are shown.

A.15 SM Prefit



(a) SM SLT (b) SM SLT

(c) SM LTT (d) SM LTT

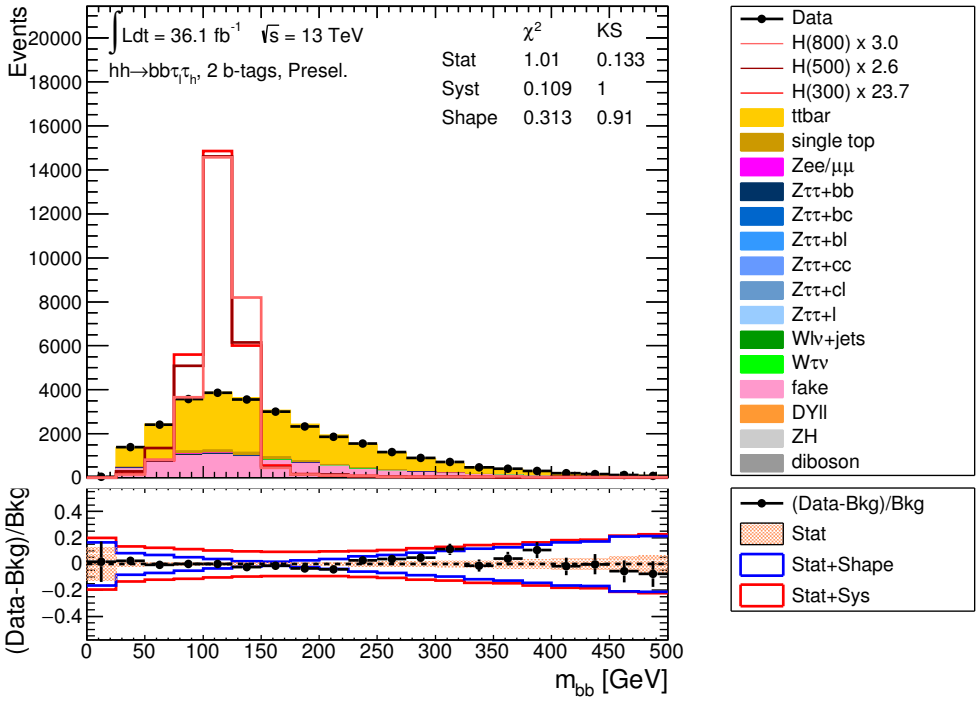
Figure A.15: Pre-fit data-MC comparisons of the BDT distributions for events at preselection with two $b-$tagged jets using arbitrary binning (left) and optimised binning (right). The SM signal and all backgrounds for the SLT (top) and LTT (bottom) channels are shown.

Appendix B

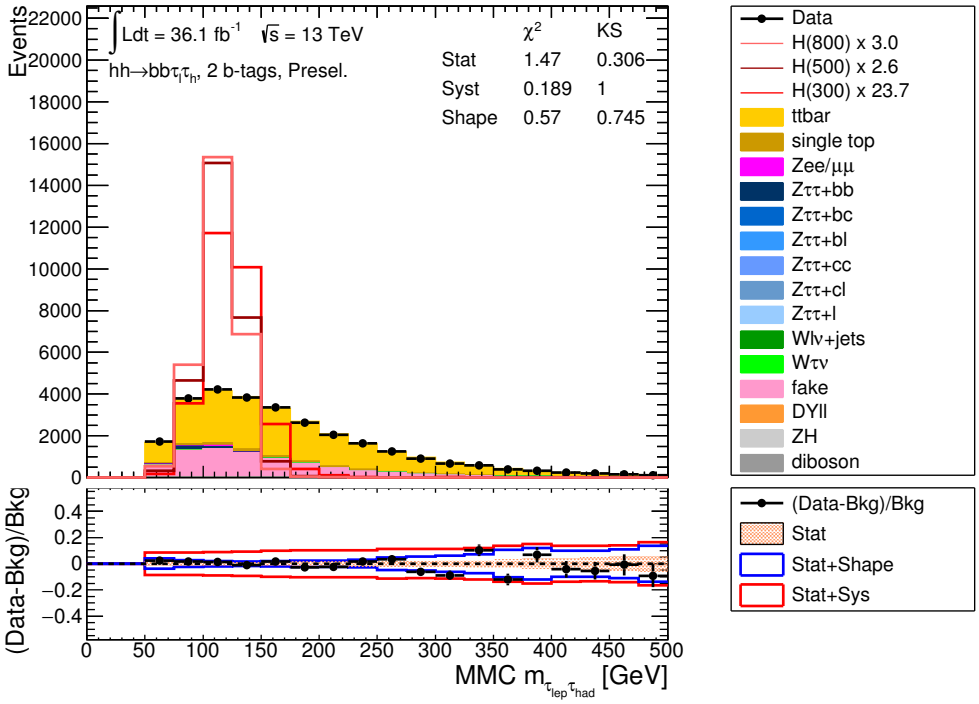
Input Variables

Data-MC comparisons of the MVA input variable distributions, described in Section 7.3.1, for the SLT channel are shown with three representative 2HDM signal mass points overlayed; 300, 500 and 800 GeV.

Figure B.1 shows the m_{bb} and $m_{\tau_\ell \tau_h}^{\mathrm{MMC}}$ distributions.
Figure B.2 shows the m_{hh} and $\Delta\phi\,(h,h)$ distributions.
Figure B.3 shows the $\Delta R\,(l,\tau)$ and $\Delta R\,(b_1,b_2)$ distributions.
Figure B.4 shows the $\Delta p_{\mathrm{T}}\,(\ell,\tau)$ and m_{T}^{W} distributions.
Figure B.5 shows the $E_{\mathrm{T}}^{\mathrm{miss}}$ and $E_{\mathrm{T}}^{\mathrm{miss}}$-$\phi$-Centrality distributions.
Figure B.6 shows the $p_{\mathrm{T}}^{b_2}$ distribution.

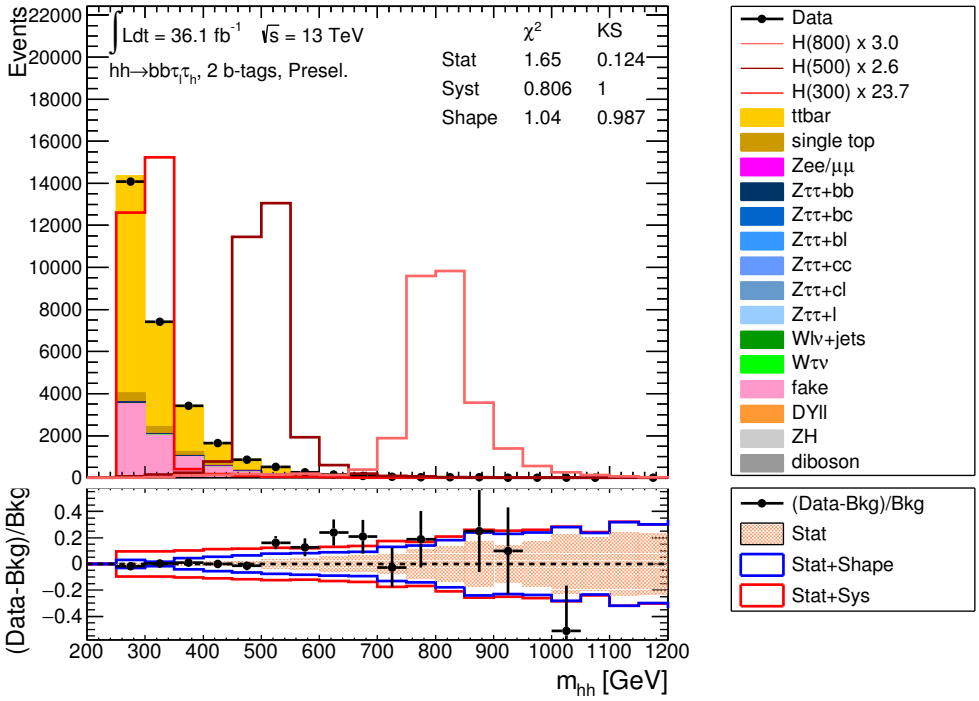


(a)

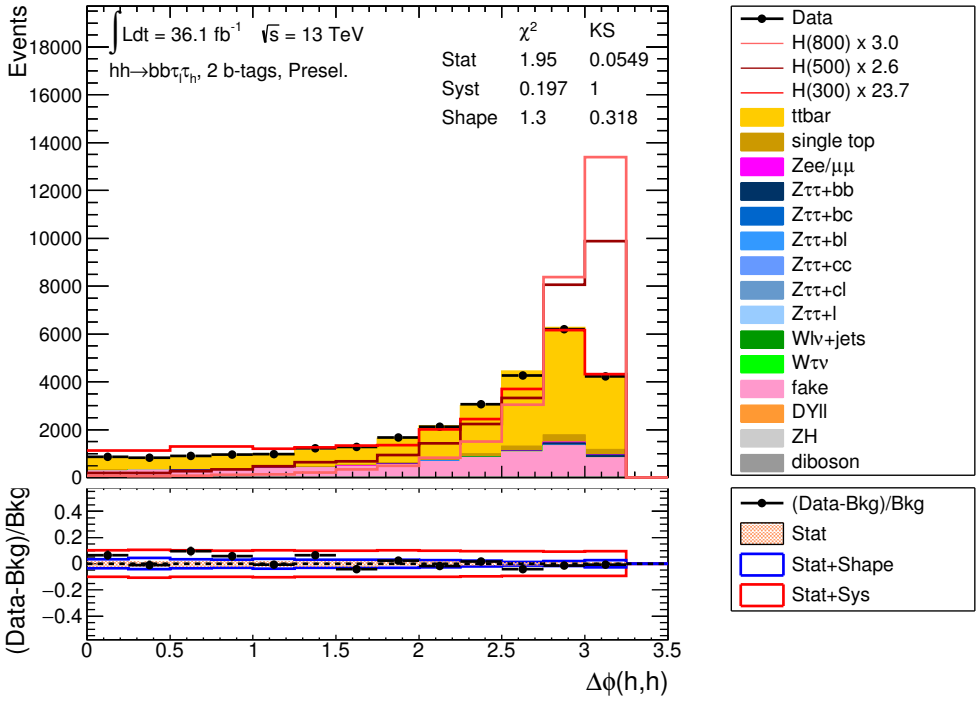


(b)

Figure B.1: Data-MC comparisons of events at preselection with two $b-$tagged jets of the discriminating variables considered for use in the MVA: (a) m_{bb}, (b) $m^{\mathrm{MMC}}_{\tau_\ell \tau_h}$.

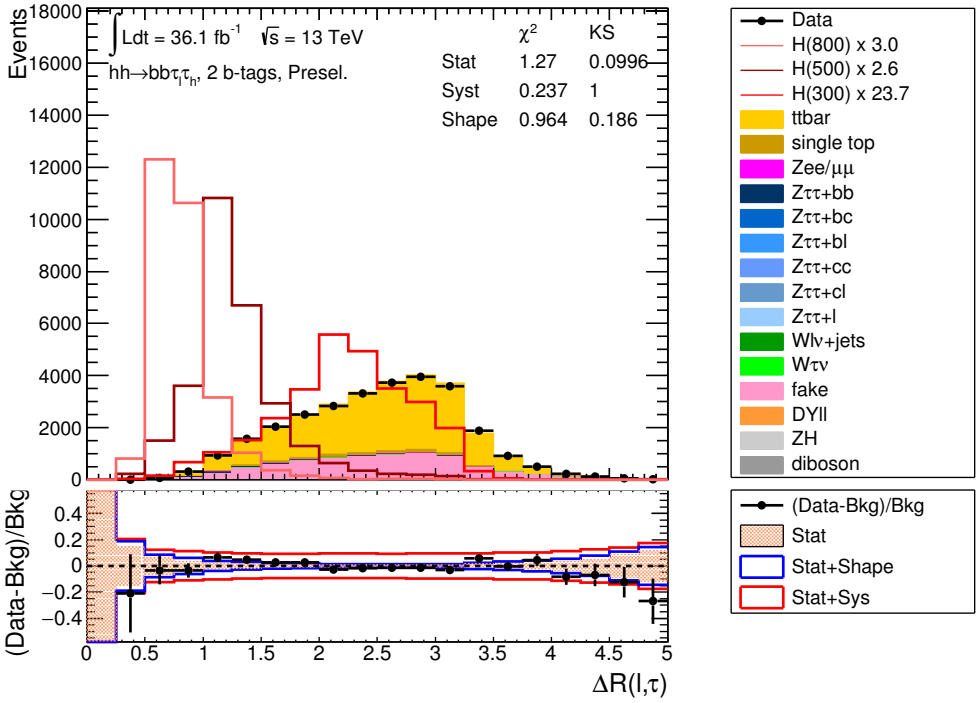


(a)

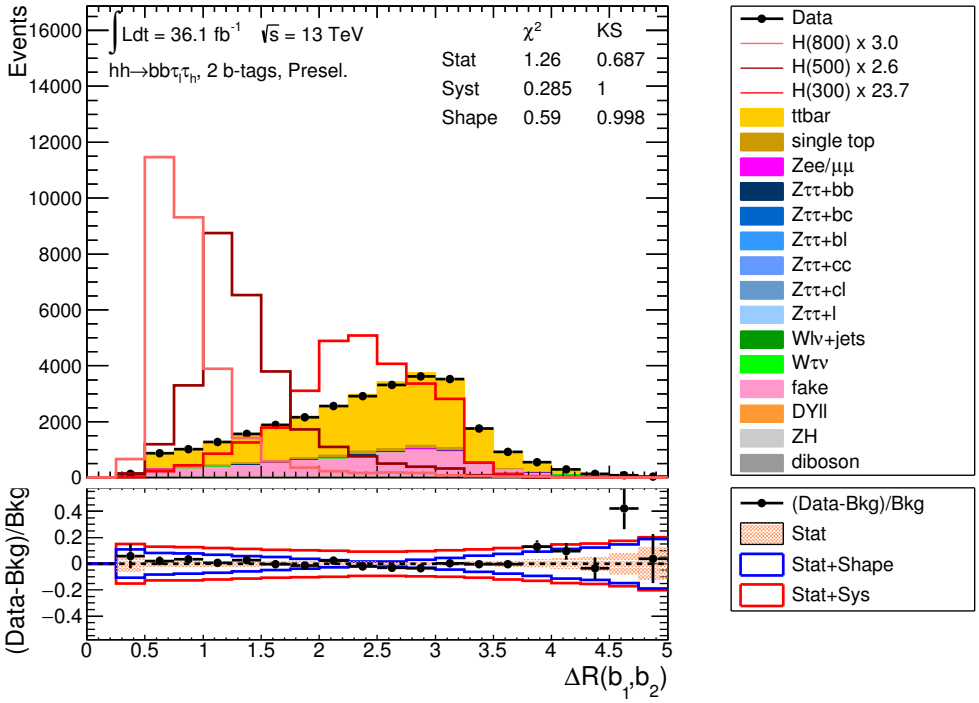


(b)

Figure B.2: Data-MC comparisons of events at preselection with two $b-$tagged jets of the discriminating variables considered for use in the MVA: (a) m_{hh}, (b) $\Delta\phi\,(h,h)$.

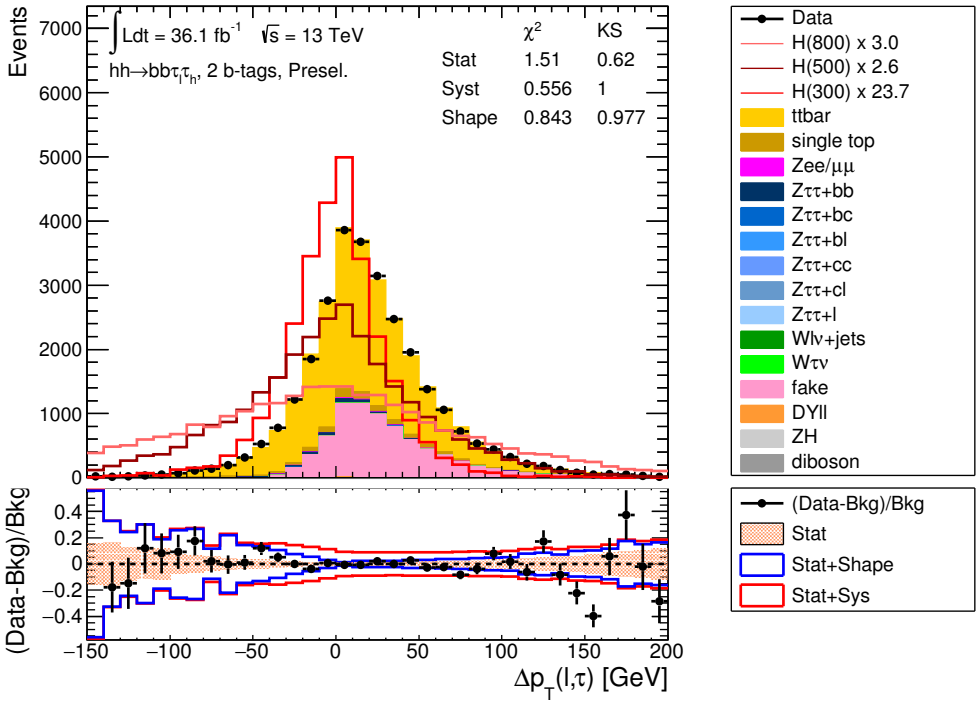


(a)

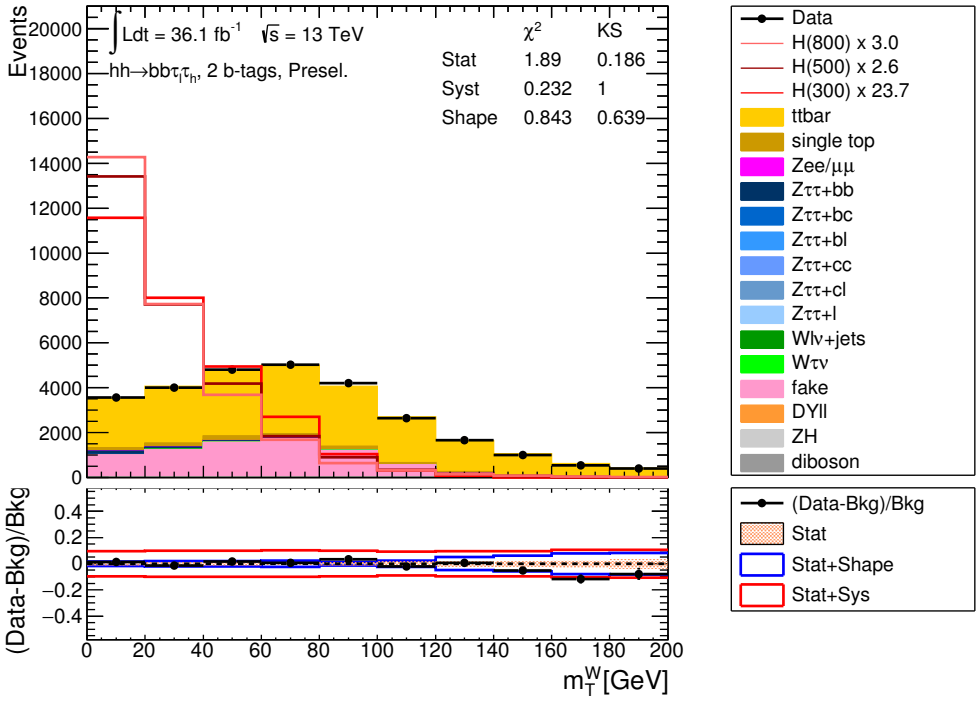


(b)

Figure B.3: Data-MC comparisons of events at preselection with two $b-$tagged jets of the discriminating variables considered for use in the MVA: (a) $\Delta R\,(l,\tau)$, (b) $\Delta R\,(b_1,b_2)$.

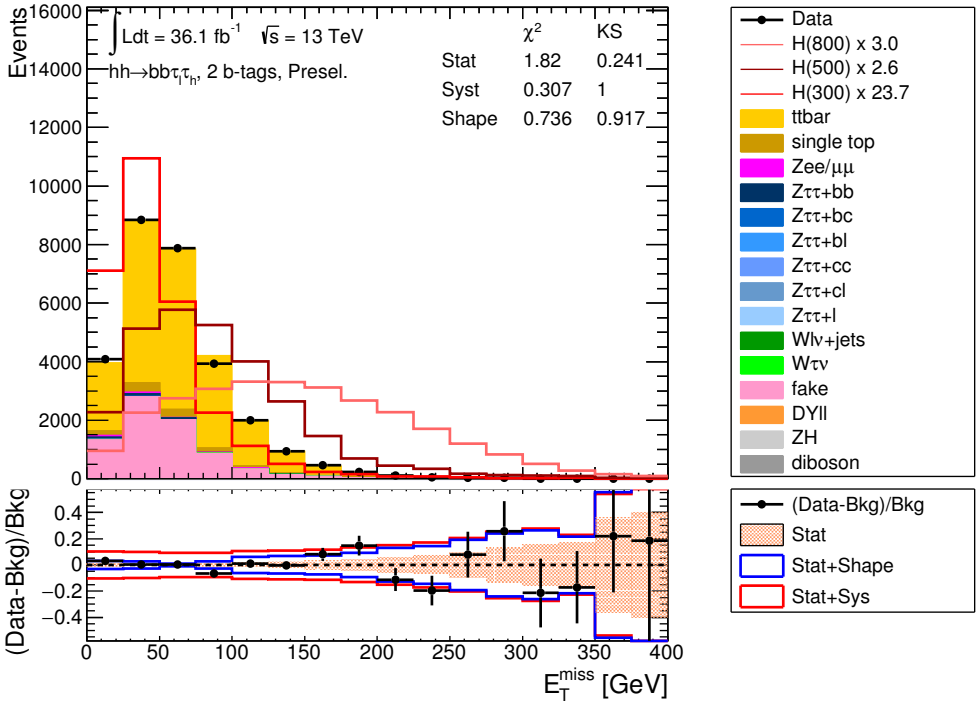


(a)

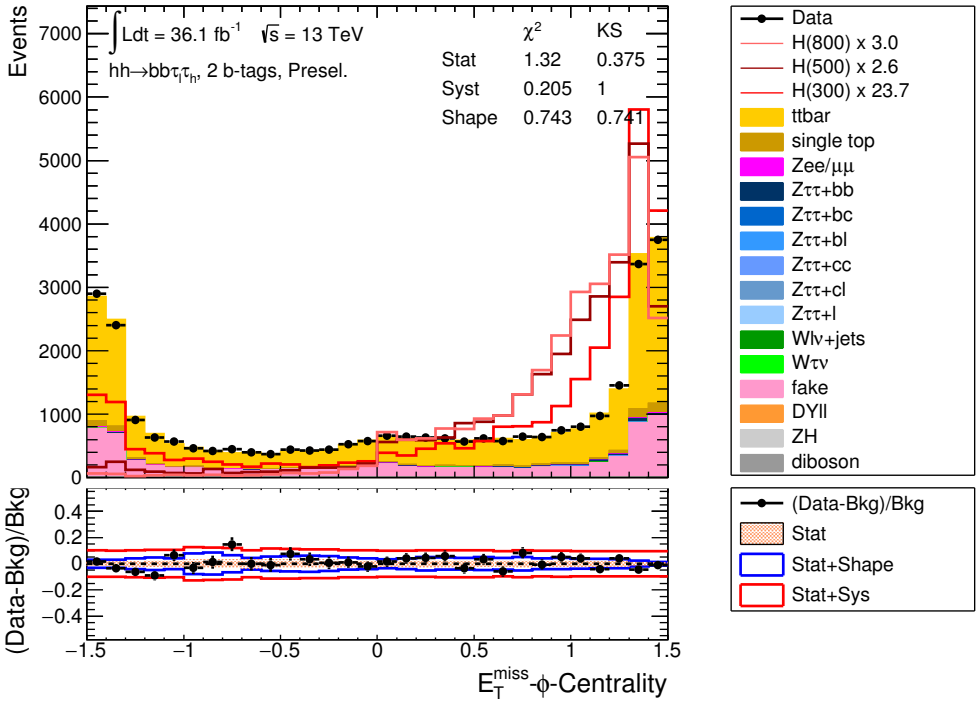


(b)

Figure B.4: Data-MC comparisons of events at preselection with two $b-$tagged jets of the discriminating variables considered for use in the MVA: (a) $\Delta p_{\mathrm{T}}\,(\ell,\tau)$, (b) m_{T}^{W}.



(a)



(b)

Figure B.5: Data-MC comparisons of events at preselection with two $b-$tagged jets of the discriminating variables considered for use in the MVA: (a) $E_{\mathrm{T}}^{\mathrm{miss}}$, (b) $E_{\mathrm{T}}^{\mathrm{miss}}$-$\phi$-Centrality.

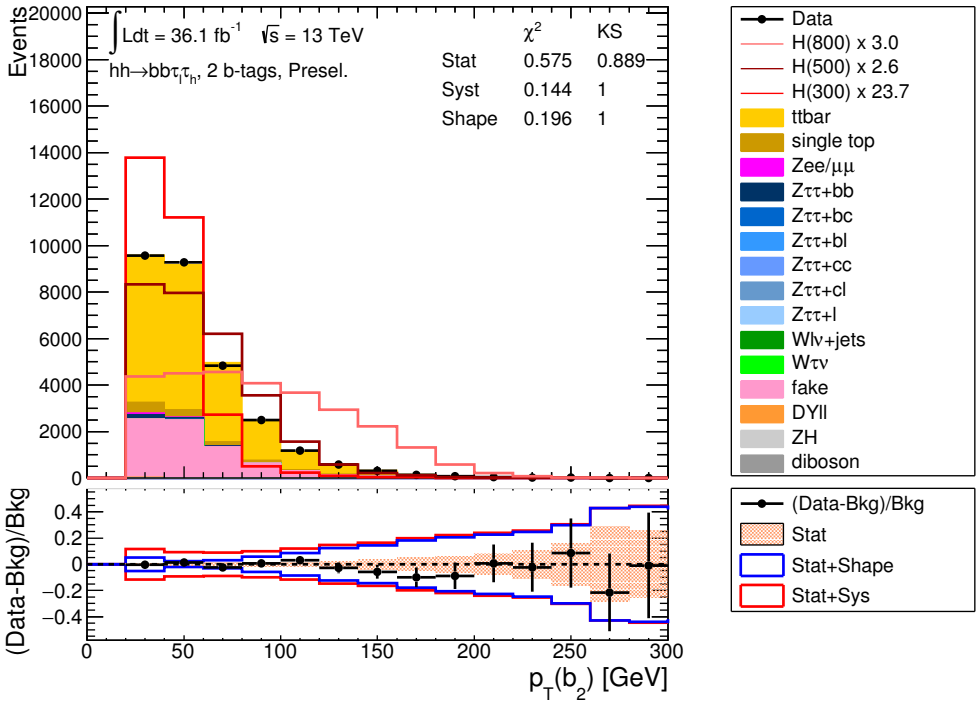


Figure B.6: Data-MC comparisons of events at preselection with two $b-$tagged jets of the discriminating variables considered for use in the MVA: $p_{\mathrm{T}}^{b_2}$.

Appendix C

SVM Optimised Hyper-Parameters

In this Appendix the optimised hyper-parameters for the SVMs presented in Chapter 11 are presented. The hyper-parameters for the multi-Gaussian SVM trained to separate the non-resonant SM signal from the $t\bar{t}$ ($Z\tau\tau + \text{HF}$) background are shown in Table C.1(C.2). The hyper-parameters for the multi-Gaussian SVMs trained to separate the $t\bar{t}$ background from the 2HDM signal for all mass points are given in Table C.3.

SM	$\Gamma_{\Delta R(b,b)}$	$\Gamma_{\Delta R(l,\tau)}$	$\Gamma_{m_{bb}}$	$\Gamma_{m_{hh}}$	$\Gamma_{m_{\tau_\ell\tau_h}^{\text{vis}}}$	$\Gamma_{m_{\text{T}}^{W}}$	C
$t\bar{t}$	0.526	0.706	0.788	0.700	0.021	0.999	10

Table C.1: Optimised hyper-parameters for the multi-Gaussian SVM trained to separate $t\bar{t}$ background and non-resonant SM signal in the LTT channel.

SM	$\Gamma_{\Delta R(b,b)}$	$\Gamma_{\Delta R(l,\tau)}$	$\Gamma_{m_{bb}}$	$\Gamma_{m_{hh}}$	$\Gamma_{m_{\tau_\ell\tau_h}^{\text{MMC}}}$	$\Gamma_{p_{\text{T}}^{bb}}$	C
$Z\tau\tau + \text{HF}$	0.994	0.985	0.859	0.993	0.865	0.852	10

Table C.2: Optimised hyper-parameters for the multi-Gaussian SVM trained to separate $Z\tau\tau + \text{HF}$ background and non-resonant SM signal in the LTT channel.

	$\Gamma_{\Delta R(b,b)}$	$\Gamma_{\Delta R(l,\tau)}$	$\Gamma_{m_{bb}}$	$\Gamma_{m_{hh}}$	$\Gamma_{m_{\tau_\ell\tau_h}^{\mathrm{MMC}}}$	$\Gamma_{p_{\mathrm{T}}^{bb}}$	C
H 260 GeV	0.502	0.499	0.505	0.497	0.500	0.504	14.1
H 275 GeV	0.479	0.400	0.995	0.932	0.932	0.307	14.9
H 300 GeV	0.206	0.705	0.997	0.896	0.999	0.987	12.5
H 325 GeV	0.070	0.662	0.999	0.497	0.548	0.576	12.5
H 350 GeV	0.879	0.858	0.999	0.874	0.776	0.600	12.9
H 400 GeV	0.520	0.519	0.997	0.982	0.888	0.647	12.5
H 450 GeV	0.995	0.989	0.998	0.811	0.958	0.997	14.9
H 500 GeV	0.983	0.386	0.795	0.543	0.860	0.787	12.9
H 550 GeV	0.430	0.279	0.487	0.978	0.510	0.968	14.5
H 600 GeV	0.861	0.010	0.938	0.368	0.116	0.738	13.4
H 700 GeV	0.020	0.486	0.939	0.990	0.532	0.602	14.0
H 800 GeV	0.010	0.660	0.636	0.598	0.452	0.999	12.6

Table C.3: Optimised hyper-parameters for the multi-Gaussian SVM trained to separate $t\bar{t}$ background and 2HDM signal in the LTT channel.